Unorthodox theoretical methods

by

Sean Alexander Nedd

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Physical Chemistry

Program of Study Committee:

Mark S. Gordon, Major Professor

Andreja Bakac

James Evans

Marek Pruski

Theresa Windus

Iowa State University

Ames, Iowa

2012

Table of Contents

Chapter 3 USING A REACTIVE FORCE FIELD IN SIMOMM TO INVESTIGATE

THE STRUCTURES OF CATALYTIC ANALOGUES ON THE PORE SURFACE

OF MESOPOROUS SILICA NANOPARTICLE SYSTEMS

Chapter 4 INCORPORATING A COMPLETELY RENORMALIZED COUPLED

CLUSTER APPROACH INTO A COMPOSITE METHOD FOR

THERMODYNAMIC PROPERTIES AND REACTION PATHS

Chapter 5   GENERAL CONCLUSIONS

Acknowledgements

First, I would like to show much thanks to Professor Mark Gordon, my advisor and mentor who has exhibited much guidance patience in all aspects of my research and my acclimatization to my life as a graduate student.

I also thank my parents, brother and family who have given so much encouragement and words of wisdom in my decision to enter into a PhD program and in how to get through such a program.

Special thanks to Yanyan Hu, who has always been a constant ear of support in all aspects of my life, particularly through my PhD process.

My friends Bosiljka Njegic, Aditya Rawal, Aaron West, Beth Prince and Ben Prince have made my time a lot more bearable through socializing and valuable advice in my research.

I have been proud to be a part of a very dynamic and supportive research group headed by Dr. Gordon. Dr. Adri Van duin and Dr. Nathan De Yonker were integral in the progress of my research, through supplementary materials and initial program codes. Dr. Victor Lin had a strong part in setting up my research and his efforts are appreciated.

Chapter 1
GENERAL INTRODUCTION

1.1 General Overview

Quantum chemistry employs techniques in mathematics, physics and chemistry to understand the behavior and properties of atoms and molecular systems. As such, quantum chemistry can be employed within many fields in the biological and physical sciences. Quantum chemistry is particularly useful in the corroboration of experimental data, as well as predicting properties that cannot be obtained via experimental means. The methods and techniques used in quantum chemistry vary in terms of the level of accuracy, as well as the amount of computational cost involved. The difficulty in quantum chemistry is in finding a low computational cost method that can offer a reasonable level of chemical accuracy for the system analyzed.

Usually, lower computational cost methods take advantage of approximations, and many of these methods are used to determine structures and properties of large systems such as biological systems. The more accurate and computationally intensive methods tend to be used to obtain a detailed understanding of the properties of small molecular systems, with a small number of atoms. This dissertation takes into account hybrid methods and force fields that attempt to maintain a low computational cost, while improving the accuracy of results on bulk systems and reaction pathways.

1.2 Dissertation Organization

This dissertation contains three chapters, the topics of which are either published or in preparation for publication.

Chapter 2 considers the use of the ReaxFF force field to correlate with NMR mobilities of amine catalytic substituents on a mesoporous silica nanosphere surface. Chapter 3 discusses the interfacing of the ReaxFF force field within the Surface Integrated Molecular Orbital/Molecular Mechanics (SIMOMM) method, in order to replicate earlier SIMOMM published data and to compare with the ReaxFF data from Chapter 2. Chapter 4 presents the development of a new correlation consistent Composite Approach (ccCA), which incorporates the completely renormalized coupled cluster method with singles, doubles and

non-iterative triples corrections towards the determination of heats of formations and reaction pathways which contain biradical species.

1.3 Theoretical Background

1.3.1 Quantum Mechanics Introduction

Quantum mechanics is applied within quantum chemistry in order to mathematically solve problems in chemistry. Quantum mechanics translates its formulae along with statistical mechanics into observable quantities that are representative of chemical properties, including thermodynamics, spectra, barriers, kinetics, and molecular properties. All of these observables originate in providing an understanding and an accurate description of the structural makeup of chemical systems.

Classically, Newtonian mechanics can be used to describe large systems, however, quantum mechanics must be used to describe non-classical particles such as electrons. The base equation that is used to understand chemical systems is the Time Dependent Schrödinger Equation (TDSE),

$$H\Psi(r,t) = i\frac{\partial \Psi(r,t)}{\partial t} \tag{1}$$

where $\Psi(r,t)$ is the wavefunction that describes the chemical system and 'H' is the Hamiltonian or energy operator on that system. The Hamiltonian is broken up into the kinetic (T) and potential (V) energy operators in,

$$H = T + V \tag{2}$$

The TDSE can take advantage of the separation of the variables position (r) and time (t). The Time Independent Schrödinger Equation (TISE) is:

$$H\Psi(r) = E\Psi(r). \tag{3}$$

The TISE is an eigenvalue equation that gives the energy eigenvalue E, and the wavefunction may be written as,

$$\Psi(r,t) = \Psi(r)e^{-Et}. \tag{4}$$

The Hamiltonian consists of energy operators that describe a particular chemical system of interest in terms of protons, electrons and their interactions. Thus, the Hamiltonian for the total energy (H_{tot}) of a chemical system can be written as

$$H_{tot} = H_e + T_n, \tag{5}$$

where T_n is the nuclear kinetic energy operator expanded to

$$T_n = -\sum_{i=1}^{N} \frac{1}{2M_i} \nabla_i^2; \nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right), \tag{6}$$

where N is the total number of atoms and M is the atomic mass of a specific nucleus. The electronic Hamiltonian (H_e) is expanded as

$$H_e = T_e + V_{ne} + V_{ee} + V_{nn}, \tag{7}$$

with the terms electronic kinetic energy (T_e), nuclear-electron potential energy (V_{ne}), electron-electron potential energy (V_{ee}), and nuclear-nuclear potential energy (V_{nn}). H_e may be expanded in the form,

$$H_e = \overbrace{-\sum_{i}^{N_{elec}} \frac{1}{2} \nabla_i^2}^{T_e} - \overbrace{\sum_{a}^{N_{nuclei}} \sum_{i}^{N_{elec}} \frac{Z_a}{|R_a - r_i|}}^{V_{ne}} + \overbrace{\sum_{i}^{N_{elec}} \sum_{j>i}^{N_{elec}} \frac{1}{|r_i - r_j|}}^{V_{ee}} + \overbrace{\sum_{a}^{N_{nuclei}} \sum_{b>a}^{N_{nuclei}} \frac{Z_a Z_b}{|R_a - R_b|}}^{V_{nn}} \tag{8}$$

N_{elec} is number of electrons, N_{nuclei} is number of nuclei, i and j are indices for electrons, a and b are indices for nuclei, r is an electron coordinate, R is the nucleus coordinate, and Z is the atomic charge.

The Hamiltonian operator used for quantum mechanical calculations is based on the Born-Oppenheimer approximation, which asserts that any change in motion of electrons is instantaneous due to the large difference in masses between protons and electrons, so there are no coupling kinetic energy terms between electrons and protons. Thus, within the Born-Oppenheimer approximation, H_e is explicitly solved at various nuclear coordinates in order to obtain a potential energy surface representing the whole chemical system. The electronic Hamiltonian is separated into one electron (h_i) and two electron (g_{ij}) terms,

$$h_i = \frac{1}{2}\nabla_i^2 + \sum_a^{N_{nuclei}} \frac{Z_a}{|R_a - r_i|} \tag{9}$$

and

$$g_{ij} = \frac{1}{r_i - r_j} \tag{10}$$

$$\sum_i h_i = T_e + V_{ne} \tag{11}$$

$$\sum_{ij} g_{ij} = V_{ee}. \tag{12}$$

Within the orbital approximation, the wavefunction, $\Psi(r)$, at a fixed set of nuclear coordinates is a product function (called the Hartree product) of all molecular orbital functions and the included antisymmetrizer giving the form,

$$\Psi(r) = A\prod_{i=0}^{N} \chi_i, \tag{13}$$

where χ is known as a spin orbital function that incorporates the spatial function (ψ) and spin (α ($+\frac{1}{2}$) or β ($-\frac{1}{2}$)), which describe the molecular orbitals for a particular chemical system such that,

$$\chi = \psi\alpha \text{ or } \psi\beta. \tag{14}$$

The problem with the Hartree product is that it does not account for indistinguishability or antisymmetry. The Slater determinant can be used to represent antisymmetry in Ψ. One Slater determinant represents one electronic configuration. A Slater determinant wavefunction (Φ_{SD}) can be expressed as,

$$\Phi_{SD} = (N!)^{-1/2} \begin{vmatrix} \chi_i(1) & \chi_j(1) & \cdots & \chi_k(1) \\ \chi_i(2) & \chi_j(2) & \cdots & \chi_k(2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(N) & \chi_j(N) & \cdots & \chi_k(N) \end{vmatrix} \tag{15}$$

where N is the number of electrons and is used to obtain the normalization factor for the determinant, where the rows hold each electron, and columns represent each spin orbital function. This Slater determinant is still based on the independent particle (orbital) model and therefore does not account for the correlation of electron-electron interactions. This limitation is addressed by electron correlation techniques, which will be discussed below.

The spatial molecular orbitals can be extended as a linear combination of atomic orbitals or LCAO in the form,

$$\psi_i = \sum_{k=0}^{N} c_k \phi_k, \tag{16}$$

where N is the number of atomic orbitals, c_k is an unknown molecular orbital coefficient, and ϕ_k is the atomic orbital/basis function. The basis functions most commonly used in quantum chemistry are Gaussian type functions GTOs [2]. GTOs have the form

$$\phi_{\zeta,n,l,m}(r,\theta,\varphi) = N Y_{l,m}(\theta,\varphi) r^{2n-2-l} e^{-\zeta r^2}, \tag{17}$$

which employ polar coordinates and spherical harmonics, or

$$\phi_{\zeta,l_x,l_y,l_z}(x,y,z) = N x^{l_x} y^{l_y} z^{l_z} e^{-\zeta r^2}, \tag{18}$$

which employs Cartesian coordinates. The Cartesian-based GTO functional form is preferred over the spherical harmonics version since many modern programs evaluate 2-electron integrals using Cartesian coordinates. These programs can transform the Cartesian GTO into the spherical harmonics GTO, which is a faster process than evaluating the 2-electron integrals using spherical harmonics GTOs. It is noted that the Cartesian-based GTOs generate more components, for example, for d-functions spherical harmonics generate 5 function components whereas the Cartesian-based GTO generates 6 function components. The Hartree Fock equations are written in the form,

$$F_i \sum_{k=0}^{N} c_{ki} \chi_k = \varepsilon_i \sum_{k=0}^{N} c_{ki} \chi_k, \tag{19}$$

which can be rearranged to

$$FC = SC\varepsilon, \tag{20}$$

where F is the Fock matrix, C is the set of molecular coefficients, S is the overlap matrix, and ε is the energy for each set of Fock matrix elements. The Fock matrix (F_{jk}) and overlap matrix (S_{jk}) are respectively,

$$F_{jk} = \langle \chi_j | F | \chi_k \rangle, \text{ and} \tag{21}$$

$$S_{jk} = \langle \chi_j | \chi_k \rangle. \tag{22}$$

Eq. 20 must be solved iteratively, using the self-consistent field or SCF technique. In the SCF technique, first a set of guessed molecular coefficients are obtained from the two electron

calculation, which is inserted into the Fock matrix, then this matrix is diagonalized in order to form a density matrix and a new set of molecular coefficients. These new coefficients are inserted into the Fock matrix and re-diagonalized until a certain user-defined threshold is achieved with the iterative scheme, and the resulting electron density is used to solve for the total energy.

In order to incorporate electron correlation, a multi-determinant wavefunction (Ψ_{MD}) is used:

$$\Psi_{MD} = a_0 \Phi_{HF} + \sum_{i=0} a_i \Phi_i, \tag{23}$$

where Φ_i wavefunction represents different possible electron configurations and the a_i are variational coefficients. There are many multi-determinant correlation methods, including Configurational Interaction (CI), Møller Plesset perturbation theory (MP), Coupled Cluster theory (CC), and Density Functional theory (DFT), as well as multi-reference methods. The single reference methods will be expanded upon in this dissertation. [3]

1.3.2 Møller Plesset Perturbation Theory

In perturbation methods, generally known under Many Body Perturbation Theory (MBPT), a small perturbation is added to an approximate solution with the idea of improving upon that solution. This perturbation (H') is added to the reference (H_0) to give the modified Hamiltonian,

$$H = H_0 + \lambda H', \tag{24}$$

where λ is 0 or 1, depending on whether the perturbation is on (1) or off (0). The wavefunction (Ψ) and the energy (E) are also expanded in a series of perturbative corrections:

$$E = \lambda^0 E^{(0)} + \lambda^1 E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + \cdots \tag{25}$$

$$\Psi = \lambda^0 \Psi_0 + \lambda^1 \Psi_1 + \lambda^2 \Psi_2 + \lambda^3 \Psi_3 + \cdots. \tag{26}$$

The superscript (n) indicates the order of the perturbation.

Because of the zeroth order Hamiltonian that is usually chosen, the first order perturbation correction to Hartree-Fock is zero. Thus, the second order Møller Plesset energy (MP2)

correction represents the first electron correlation energy correction value for Møller Plesset perturbation theory. The final MP2 energy correction is:

$$E^{(2)} = \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{\left(\left\langle \phi_i \phi_j \middle| \phi_a \phi_b \right\rangle - \left\langle \phi_i \phi_j \middle| \phi_b \phi_a \right\rangle\right)}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}, \qquad (27)$$

with virtual orbitals a and b, and occupied orbitals i and j.

MP2 is not very computationally intensive, even though MP2 scales as M_{basis}^5, where M_{basis} is the number of basis functions. In terms of accuracy, MP2 accounts for about 80% to 90% of the correlation energy. [3]

1.3.3 Coupled Cluster Theory

The Couple Cluster (CC) theory includes all many body corrections to infinite order. In order to achieve this extension, an excitation operator (T) is used such that

$$T = T_1 + T_2 + T_3 + T_4 + \cdots + T_{N_{elec}}, \qquad (28)$$

where the T_i operator operates on a reference wavefunction in order to generate all ith excited Slater determinants. For example, a single excitation equation would be in the form of

$$T_1 \Phi_0 = \sum_i^{occ} \sum_a^{vir} t_i^a \Phi_i^a, \qquad (29)$$

and the double excitation equation would be

$$T_2 \Phi_0 = \sum_{i<j}^{occ} \sum_{a<b}^{vir} t_{ij}^{ab} \Phi_{ij}^{ab}, \qquad (30)$$

The t coefficients are called amplitudes. For example, in Eq. 29 the t coefficient is the singles excitation amplitudes, while in Eq. 30 the t coefficient is the doubles excitation amplitudes. The CC wavefunction (Ψ_{CC}) can be written as

$$\Psi_{CC} = e^T \Phi_0, \qquad (31)$$

where e^T generates the excitation operators in the form of

$$e^T = \sum_{k=0}^{\infty} \frac{1}{k!} T^k, \qquad (32)$$

The Schrödinger equation can be re-written using the CC operator as

$$He^{T}\Phi_{0} = Ee^{T}\Phi_{0}, \tag{33}$$

The CC energy expression for just singles and doubles excitations (CCSD) can be written as:

$$E_{CCSD} = E_0 + \sum_{i<j}^{occ}\sum_{a<b}^{vir}\left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a\right)\left(\langle\phi_i\phi_j|\phi_a\phi_b\rangle - \langle\phi_i\phi_j|\phi_b\phi_a\rangle\right). \tag{34}$$

In Chapter 4, the CC methods of note are the coupled cluster method with singles, doubles and perturbative triples (CCSD(T)) [10], and the left-eigenstate completely renormalized coupled cluster method with singles doubles, and non-iterative triples corrections (CR-CC(2,3)) [11].

1.3.4 Density Functional Theory

There are several advantages of using the density function over the wavefunction. The density function is simpler since there are only 3 variables, whereas wavefunction based methods depend on 3N variables, where is N is the number of electrons. Density Functional Theory (DFT) avoids the use of the wavefunction by using the density function in order to obtain physical observables.

The DFT method relies on the Kohn-Sham equations that are analogous to the Hartree-Fock equations:

$$\left[-\frac{\nabla^2}{2} + V_{KS}[\rho](\vec{r})\right]\phi_i(\vec{r}) = \varepsilon_i\phi_i(\vec{r}), \tag{35}$$

where
$$\rho(\vec{r}) = \sum_i^{occ}|\phi_i(\vec{r})|^2, \tag{36}$$

$$V_{KS}[\rho](\vec{r}) = V_{ext}(\vec{r}) + V_H[\rho](\vec{r}) + V_{xc}[\rho](\vec{r}), \tag{37}$$

$$V_{ext}(\vec{r}) = \sum_a V_a(\vec{r} - \vec{R}_a), \tag{38}$$

$$V_H[\rho](\vec{r}) = \int d\vec{r}' \frac{\rho(\vec{r}')}{|\vec{r}-\vec{r}'|}, \tag{39}$$

and
$$V_{xc}[\rho](\vec{r}) = \frac{\delta E_{xc}[\rho]}{\delta\rho(\vec{r})}. \tag{40}$$

V_{KS} is the Kohn-Sham potential and has a functional dependence on the electron density. The Kohn-Sham equations are solved iteratively, in a similar manner to the HF equations. Both

the Hartree potential (V_H), and the exchange-correlation potential (V_{xc}) have a functional dependence on the electron density, however, V_{xc} is not known and is approximated. [12]

The main classes of density functional approximations are the Local Density Approximation (LDA), the General Gradient Approximation (GGA), the meta General Gradient Approximation (meta-GGA), and the hybrid General Gradient Approximation (hybrid-GGA).

In LDA, the local density (density in a small neighborhood of a spatial point) is treated as a uniform electron gas, which is based on the assumption that the density varies slowly. If spin polarization is added due to treating the alpha and beta electron contributions separately, the method is termed LSDA or Local Spin Density Approximation and is the typically used version of LDA.

In GGA, the local density is described both by its value and its gradient of the electron density. In the meta-GGA methods the kinetic energy density (equivalent to the Laplacian of the electron density) is added to the description of the local density by GGA.

In the hybrid-GGA methods the Hartree-Fock exchange energy is added to the exchange-correlation energy. The B3LYP [13] method, which is based on the Becke 3 parameter exchange energy functional and the Lee, Yang and Parr (LYP) correlation energy functional is a typically used density functional approximation and is a hybrid-GGA method. Another such hybrid-GGA method is the PBE0 [14] method, which is an improvement on the GGA method of PBE (Perdew-Burke-Ernzerhof). B3LYP and PBE0 are noted specifically since they are used as the DFT methods of choice in Chapter 3. [3]

Overall, DFT has a computational cost similar to that of Hartree Fock theory, but with a higher accuracy of computed results.

1.3.5 Composite Methods

Composite methods combine energy corrections from lower (less accuracy and lower computational cost) QM methods (LM) and lower basis sets (LB) in order to provide more accurate final energies representing higher (more accuracy and higher computational cost) level methods (HM) and higher basis sets (HB) in the form of E(HM/HB). In terms of these energy corrections, this final energy might be expressed in the form:

$$E(HM/HB) = E(LM/HB) + \left(E(HM/LB) - E(LM-LB)\right), \tag{41}$$

where *E(LM/HB)* is the reference energy of the molecular system, while the other energy calculations are single point energy calculations. (*E(HM/LB)* – *E(LM/LB)*) is the energy correction that accounts for the HM. There are several versions of Composite methods such as Gaussian and Weizmann methods, which are highlighted in Ch. 4. However, one of the composite methods known as the correlation consistent Composite Approach using the S4 complete basis set extrapolation (ccCA-S4), which is used in Ch. 4, will be expanded upon here in order to further illustrate the general setup of composite methods by showing that many energy correction terms may be used to obtain a final composite method energy value.

As indicated and further detailed in Ch. 4, ccCA-S4 [15] uses B3LYP/6-31G(2df,p) for structural optimizations and Hessian calculations. For single point energy calculations ccCA-S4 employs MP2/aug-cc-pVQZ, MP2/aug-cc-pVTZ, MP2/cc-pVTZ-DK (Douglas-Kroll-Hess scalar relativistic effects), CCSD(T)/cc-pVTZ (coupled cluster with singles, doubles and non-iterative triples method), MP2/aug-cc-pCVTZ (core-valence basis set), and MP2/cc-pVTZ calculations. Each of these calculations is used towards a type of energy correction, which will now be outlined. There are four corrections to consider. The first deals with a complete basis set two point extrapolation in which the calculations of MP2/aug-cc-pVQZ and MP2/aug-cc-pVTZ give the two energy values required for the extrapolation to give an energy called MP2/CBS. The other corrections are for the Douglas-Kroll scalar relativistic effects ($\Delta E(SR-MP2)$), coupled cluster beyond MP2 ($\Delta E(CC)$), and core-valence relativistic effects ($\Delta E(CV)$). Using the above single point energy calculations the last three corrections are written as,

$$\Delta E(MP2 - SR) = (MP2/cc - pVTZ - DK) - (MP2/cc - pVTZ), \tag{42}$$

$$\Delta E(CC) = E(CCSD(T)/cc - pVTZ) - E(MP2/cc - pVTZ), \tag{43}$$

$$\Delta E(CV) = E(MP2(full)/aug - cc - pCVTZ) - E(MP2/aug - cc - pVTZ), \tag{44}$$

where MP2(full) indicates that all electrons are used in the MP2 calculation.

1.3.6 Force Fields

Force fields (or molecular mechanics) calculate potential energy surfaces by using parametrized nuclear coordinate functions, which are fitted to experimental data or some higher-level computational data. Essentially, force fields use atom positions, which gives a significant decrease in computational effort when compared to computational methods that depend on the number of electrons in a system. Thus, force fields are represented as model potentials and are divided into energy corrections that describe a chemical system. These energy corrections are obtained using classical mechanical equations. The final force field energy (E_{FF}) is divided into bonded energy corrections and non-bonded energy corrections. The typical E_{FF} is set up as follows,

$$E_{FF} = E_{Str} + E_{Bend} + E_{Tors} + E_{vdW} + E_{El} + E_{Cross} \tag{45}$$

where E_{Str}, E_{Bend}, E_{Tors}, E_{vdW}, E_{el}, and E_{Cross} represent energy corrections due to bond stretching, bond bending, bond torsions, Van der Waals, electrostatics, and coupling terms (between E_{Str}, E_{Bend}, and E_{Tors}) respectively. E_{VdW} and E_{el} are the non-bonded energy corrections; other non-bonded corrections may be added in certain force field methods. [3] There are several types of force fields as laid out in Chapter 2. In this dissertation, the Reactive Force Field (ReaxFF), which was developed by William Goddard III and Adri van Duin, is used. ReaxFF is unique in that certain functions that account for bond orders are added in order to allow the force field better describe bond breaking and formation, which is required for obtaining more accurate chemical properties. Most force fields use fixed parameters and fixed bond connectivity that make it difficult for those force fields to obtain accurate chemical properties that come close to quantum mechanical calculations, however, ReaxFF is gives chemical properties similar to quantum mechanics. The ReaxFF energy (E_{System}) is written as

$$\begin{aligned} E_{System} = E_{Bond} + E_{over} + E_{under} + E_{lp} + E_{val} + \\ E_{pen} + E_{tors} + E_{conj} + E_{vdW} + E_{Coulomb} + E_{H-bond} \end{aligned}, \tag{46}$$

where E_{Bond} is a bond energy term that is separated into separate forms related to the varying bond order within a chemical system. The bond order term is expressed in a separate functional form that is used in other ReaxFF energy correction terms. E_{over} is the over-coordination energy terms that accounts for coordination numbers larger than the predicted

coordination numbers based on valence bond theory. E_{under} is the under-coordination energy term and is the converse to the over-coordination term. E_{lp} is the lone pair energy term due to the presence of lone pairs. E_{val} is valence energy term obtained from valence angles where the bond order function is used. E_{pen} is the energy penalty term that corrects for specific systems that have stability problems with two double bonds. E_{tors} is the torsion angle energy term, which depends on the bond order function. E_{conj} is the energy term that accounts for conjugated systems.

The last three terms account for non-bonded interactions. E_{vdW} is the van der Waals energy correction term, which depends on a distance corrected Morse-potential. $E_{Coulomb}$ is the coulomb energy correction term determined using the Electron Equilibration Method; the ReaxFF method allows the charges to vary, which is important in obtaining more accurate electrostatics. $E_{H\text{-}Bond}$ is the hydrogen bonding energy term that is bond order dependent term. Other terms have been added to the list of bonded interaction terms to account for special system interactions including for example, three-body conjugation, four-body conjugation, and triple bond corrections.

The ReaxFF method uses parameters developed for specific chemical systems, such as for hydrocarbon, silicon and silica systems. These parameters, based on a large number of chemical systems in a benchmark set, are obtained using either experimental data or DFT-based data. The benchmark structures allow ReaxFF to account for many possible types of interactions found in these structures, and in Chapter 2 and 3 the silicon/silica parameter set is used to analyze a silica system. The ReaxFF energies obtained are more accurate than typical force field techniques, while maintaining the low computational cost associated with all molecular mechanics methods.

1.3.7 Quantum Mechanical/Molecular Mechanical Methods

Quantum Mechanical/ Molecular Mechanics (QM/MM) methods are hybrid computational methods. QM/MM methods are used to obtain relative energies of systems that have a high computational cost if the system were to be analyzed purely quantum mechanically. These QM/MM methods isolate an active site, known as a reactive site model (RSM) that represents an area of chemical importance and is found within typically a large

molecular system. A QM method is chosen to obtain a user-defined level of accuracy on the RSM, however, a lower level method or MM method that has less computational cost and is usually less accurate than the QM method, is used to describe the remaining section or bulk model (BM) of the molecular system. A force field is used as the MM method. Thus, QM/MM methods have the benefit of accurately describing a RSM, while reasonably describing the BM all at low computational cost.

Within a QM/MM structural optimization calculation, the RSM is separated from the BM where certain atoms in the zone of separation are capped by hydrogen atoms. The RSM is then optimized by the QM method, followed by a relaxation of the BM coordinates by the MM method. The gradients used for the QM calculation is then adjusted as they are affected by the relaxation of the BM, then the RSM is optimized again. This optimization scheme is iterated until some convergence criteria are met. The QM/MM method used in this dissertation is known as Surface Integrated Molecular Orbital/Molecular Mechanics or SIMOMM and will be expanded upon in Ch. 3.

1.3.8 Computational Programs

All coding and calculations was done using the General Atomic and Molecular Electronic Structure System (GAMESS) [16]. GAMESS is a computational package that incorporates quantum mechanical, molecular mechanical and hybrid methods. The ReaxFF code from Dr. Adri van Duin's group as outlined in Chapter 2, was incorporated into GAMESS and was also manipulated for the calculation of SIMOMM calculations as mentioned above. The Composite method code was also implemented into GAMESS.

1.4 References

[1] J. C. Slater, Phys. Rev. **36**, 57 (1930).

[2] S. F. Boys, Proc. R. Soc. (London) A **200**, 542 (1950).

[3] F. Jensen, *Introduction to Computational Chemistry*, 2 ed. (Wiley, Odense, 2007).

[4] C. C. J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).

[5] G. G. Hall, Proc. R. Soc. (London) **A205**, 541 (1951).

[6]I. N. Levine, *Quantum Chemistry*, 5th ed. (Prentice-Hall Inc., Upper Saddle River, New Jersey, 2000).

[7]C. D. Sherrill and H. F. Schaefer, Adv. Quant. Chem. **34**, 143 (1999).

[8]C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).

[9]A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. (Dover Publications, Inc., Mineola, New York, 1989).

[10]G. E. Scuseria and T. J. Lee, J. Chem. Phys. **93**, 5851 (1990).

[11]P. Piecuch and M. Wloch, J. Chem. Phys. **123**, 224105 (2005).

[12]N. Fernando, A. Castro, and M. A. L. Marques, Lecture Notes in Physics **620**, 218 (2003).

[13]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[14]M. Ernzerhof and G. E. Scuseria, J. Chem. Phys. **110**, 5029 (1999).

[15]N. J. De Yonker, T. R. Cundari, and A. K. Wilson, J. Chem. Phys. **124**, 114104 (2006).

[16]M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, and N. Matsunaga, Journal of Computational Chemistry **14**, 1347 (1993).

Chapter 2

USING A REACTIVE FORCE FIELD TO CORRELATE MOBILITIES OBTAINED FROM SOLID-STATE ^{13}C NMR ON MESOPOROUS SILICA NANOPARTICLE SYSTEMS

A paper published in the Journal of Physical Chemistry C

Sean A. Nedd, Takeshi Kobayashi, Chih-Hsiang Tsai, Igor I. Slowin, Marek Pruski, and

Mark S. Gordon

2.1 Abstract

Theoretical calculations and solid-state ^{13}C NMR have been used to determine the conformation, relative energies and behavior of organic functional groups covalently bound within the pores of mesoporous silica nanoparticles (MSNs). The calculations were performed using the ReaxFF reactive force field for model surfaces consisting of a four-layer silica slab with one or two functional groups: N-(2-aminoethyl)-3-aminopropyl- (AAP), N-[N-(2-aminoethyl)-2-aminoethyl]-3-aminopropyl- (AEP) or 3-cyanopropyl- (CP). The results indicate that the AAP and AEP groups exist primarily in the prone orientation, while CP can almost equally occupy both the prone and upright orientations in CP-MSN. This is in agreement with the solid-state ^{13}C NMR experiments, which suggest that the AAP and AEP functionalities remain rigid on the NMR time scale (in this case sub-millisecond), whereas the CP substituent executes faster motions. These conformations are most likely governed by the hydrogen bonds between the amine moieties of the functional groups and the silanol groups on the silica surface. ReaxFF can be used to study a system, which requires a large-scale model, such as the surface of an organo-functionalized heterogeneous catalyst, with

higher accuracy than the conventional MM and at a lower computational cost than ab initio, quantum mechanical calculations.

2.2 Introduction.

Mesoporous silica nanospheres (MSNs), such as MCM-41, have shown promising potential for applications in heterogeneous catalysis due to their large surface area, well defined pore structure and narrow pore size distribution.[1-3] The MSN materials in which catalytic substituents are immobilized within the inner surfaces have been widely studied,[3-6] however, there is currently little understanding of the conformation and behavior of the catalytic functionalities within the pores. In this study, three functional substituents, N-(2-aminoethyl)-3-aminopropyl- (AAP), N-[N-(2-aminoethyl)-2-aminoethyl]-3-aminopropyl-(AEP) and 3-cyanopropyl- (CP) (Figure 1), were covalently attached onto the inner pore surface of the MSNs, and their conformations and mobilities were studied using theoretical calculations and validated by comparison with solid-state NMR experiments.

Solid-state NMR can be employed to determine the structure and mobility of catalytic substituents attached to the MSN surface. For example, the dynamics of the ^{1}H-^{13}C cross polarization process, described by the time constant T_{CH}, depends on the effective ^{1}H-^{13}C magnetic dipole-dipole interaction and thus on the atomic-level mobility. Theoretical calculations can be used to obtain energy-minimized structures and assess their stabilities.[7] The choice of computational method must take into account the size of the overall system, because MSN structures contain hundreds to thousands of atoms. The chosen method must also be able to account for the various types of interactions to be described; e.g., covalent bonds and much weaker hydrogen bonds in MSN compounds due to the presence of many

silicon atoms and hydroxyl groups. The preferred approach would be to use *ab initio* quantum mechanics (QM) to predict structures and relative energies; however, the more reliable QM methods that are normally used for large or bulk systems, such as second order perturbation theory (MP2)[8] and density functional theory (DFT)[9] require more computational resources than are typically available. The use of *ab initio* simulations on bulk silica surface systems has thus been limited to the incorporation of low computational cost (little or no polarization or diffuse functions) basis sets, which are usually insufficient for systems that contain many weak electrostatic interactions, such as silica systems.[10] This limitation also extends to *ab initio* structural optimizations of small silica structures defined within a unit cell or some periodic boundary condition; which do not fully represent bulk silica systems in MSN compounds, since these silica systems are amorphous in nature. [11-14]

A more computationally tractable alternative to QM methods is to employ model potentials, such as those incorporated in many molecular mechanics (MM) force fields. In general, MM methods (for example the Allinger MM methods[15-17] and DREIDING[18]) require orders of magnitude less computer resources (e.g., time, disk, memory) than do reliable QM methods for a given system. Most MM force fields, however, are not able to respond well to a changing molecular environment, as one would encounter in heterogeneous catalysis, or more generally in bond-breaking processes. In this context, a particularly appealing MM approach is the bond order dependent ReaxFF reactive force field.[19,20] ReaxFF is a general bond-order dependent potential that employs varied contributions to the interaction energy including bond, valence, torsion, conjugation, and under-/overcoordination (correction for total bond orders) energies, as well as Van der Waals and Coulomb interactions. ReaxFF parameters are optimized using a training set, based on experimental data as well as density

functional theory (DFT)[9] structures and energies. The parametrization, which also includes non-bonded interactions such as dispersion interactions, includes fitting to internal structural parameters (e.g., bond lengths, angles, torsions) and energy barriers for related processes within the system of interest. The ReaxFF method maintains the very low computational cost that is characteristic of MM methods, while obtaining reliable structures and related energetics. In particular, because of its flexibility that is embodied in the bond order dependence, ReaxFF can be used to study reaction mechanisms. The current paper presents a preliminary ReaxFF study of the structures and relative energies of substituted MSN prototypes.

2.3 Materials and Methods

2.3.1 Experimental.

MSN materials. The samples were prepared using a co-condensation method following previously described procedures.[21,22] Cetyltrimethylammonium bromide (CTAB), tetraethoxysilane (TEOS), and the organoalkoxysilane precursors, including (3-cyanopropyl)triethoxysilane, [N-(2-aminoethyl)-3-aminopropyl]trimethoxysilane, and N-[N-(2-aminoethyl)-2-aminoethyl]-3-aminopropyltrimethoxysilane, were purchased from Aldrich and used as received. The reaction mixture of CTAB (2.0 g, 5.49 mmol), 2.0 M of NaOH(aq) (7.0 ml, 14.0 mmol) and H_2O (480 g, 26.67 mol) were heated at 80 °C for 30 min with stirring. To the resulting clear solution, TEOS (9.34 g, 44.8 mmol) and the organoalkoxysilane precursor (5.75 mmol) were injected sequentially and rapidly, forming a white precipitate. The solution was maintained at 80 °C for 2 h under stirring at 550 rpm. The product was isolated by hot filtration, washed with copious amounts of water and methanol,

and dried under vacuum at room temperature. An acid extraction was performed in a methanol (100 ml) mixture of concentrated hydrochloric acid (1.0 ml) and as-made materials (1.0 g) at 60 °C for 6 h. The resulting surfactant-free solid products were filtered and washed with water and methanol, and dried under vacuum at room temperature. The samples containing CP, AAP and AEP functionalities are referred to as CP-MSN, AAP-MSN and AEP-MSN respectively.

Solid-state NMR. Solid-state NMR experiments were performed on a Chemagnetics Infinity 400 spectrometer, equipped with a 5-mm magic angle spinning (MAS) probe and operated at 400.0 MHz for ^{1}H, 100.6 MHz for ^{13}C and 79.5 MHz for ^{29}Si nuclei. The samples were packed in MAS zirconia rotors and spun at 10 kHz. The ^{29}Si measurements were carried out using direct polarization under MAS (DPMAS) to obtain quantitative estimates of the surface coverage of the functionalities. To enhance the sensitivity, the single-pulse excitation of ^{29}Si nuclei was followed by a Car-Purcell-Meiboom-Gill (CPMG) sequence of 10 π pulses (in 10 ms intervals), which produced a series of spin-echoes.[23, 24] The CPMG echoes were collected under two-pulse phase-modulation (TPPM) ^{1}H decoupling,[25] providing a 3-fold gain of signal without detectable spectral distortions. For each sample, 296 scans were acquired using a recycle delay of 300 s.

The ^{13}C spectra were acquired to determine the structure and the mobility of the organic functionalities. The measurements used tangently ramped cross-polarization magic angle spinning (^{13}C{^{1}H} CPMAS) to enhance the polarization of carbon nuclei and increase the repetition rate of data acquisition. The mobilities of silica-bound AAP, AEP and CP were determined by monitoring the ^{13}C signal intensity as a function of the cross-polarization contact time τ_{CP}[23] under TPPM ^{1}H decoupling. To this end, 20 spectra were acquired for each

sample by varying τ_{CP} between 5 μs and 10 ms, each requiring 4800 scans with 1.5 s recycle delay. The RF magnetic fields used in all measurements were as follows: ν_{RF}^{H} = 40 kHz, ν_{RF}^{Si} = 50 kHz and ν_{RF}^{C} = 50 kHz. ^{1}H, ^{13}C and ^{29}Si chemical shifts were referenced with respect to tetramethylsilane (TMS) at 0 ppm.

2.3.2. Computational Details.

Reactive Force Field (ReaxFF). All ReaxFF computations were done using the ReaxFF$_{SiO/Si}$ version (reactive MD-force field Si/O/C/H for polydimethylsiloxane (PDMS)).[20,27] Geometries were optimized using the conjugate gradient converging scheme and a very strong convergence criterion (0.001 kcal/mol/angstrom). The ReaxFF calculations reported here were produced by the implementation in GAMESS (General Atomic and Molecular Electronic Structure System).[28,29]

The structural optimizations were performed with no constraints for models consisting of a four-layer silica slab, referred to as 4L-MSN (Figure 2), and one or two of a given organic functional group, which represents the inner surface of functionalized MSNs. Since the local structure of amorphous silica is known to be similar to that of β-cristobalite, with the external surface being analogous to the (111) surface, the same topology was used to construct the 4L-MSN prototype with a surface area of 2.5 nm^2.[30,31] The Si-O-H bond angle was initially set at 121°, prior to geometry optimizations.[29] One functional group was attached to a site labeled P, with two initial conformations representing almost parallel (prone) and perpendicular (upright) orientations relative to the 4L-MSN surface. The final optimized organic substituent structures or functional groups were all obtained using one straight chain configuration that was optimized in the prone to upright positions in order to

allow for comparison with the NMR mobility data. The ReaxFF optimized structures obtained in this manner were used to determine the relative stabilities for the three catalytic substituents in the upright vs. the prone orientation. In order to interrogate the intermolecular interactions between neighboring functional groups, additional substituents were introduced at sites X, Y or Z located 1.15, 0.92 and 0.52 nm from P, respectively (Figure 2e). The number of interacting molecules was chosen based on the average surface concentration measured by solid-state NMR, whereas the specific locations were determined by the available sites on the model surface.

2.4 Results and Discussion.

2.4.1 Solid-state NMR. The results of ^{29}Si DPMAS (direct polarization magic angle spinning) measurements are shown in Figure 3 and Table 1. The signals centered at around -90, -100 and -110 ppm represent silicon sites Q^2 (($\equiv$SiO)$_2$**Si**(OH)$_2$), Q^3 (($\equiv$SiO)$_3$**Si**OH) and Q^4 (($\equiv$SiO)$_4$**Si**), respectively.[33] The carbon-bearing silicons T^2 (($\equiv$SiO)$_2$**Si**(OH)R) and T^3 (($\equiv$SiO)$_3$**Si**R) are known to resonate at around -55 and -65 ppm, respectively. The relative concentrations of Q^n and T^n sites were determined by deconvolution of the ^{29}Si DPMAS spectra into weighted linear combinations of Gaussian peaks (Table 1). Reliable fits were obtained without imposing any constraints on the positions and widths of individual components, as verified by numerical and visual comparison of the experimental and calculated spectra. The amounts of organic functionalities were obtained from the overall concentration ratios $(T^2+T^3)/(T^2+T^3+Q^2+Q^3+Q^4)$ and easily translated into the surface concentrations per 1 nm^2 using the BET surface areas.[21,22]

Figure 4 shows the $^{13}C\{^1H\}$ CPMAS spectra and the assignments of ^{13}C resonance signals,[21,22,34] which indicate that the MSN samples are indeed functionalized as intended and are essentially surfactant-free. The cross polarization time constants (T_{CH}) for individual carbons in the functional groups were derived from the 'build-up' of ^{13}C magnetizations during the cross-polarization period[26] (Table 2). The T_{CH} values depend on the effective ^1H-^{13}C magnetic dipole-dipole interactions and provide information about the degree of atomic-level motions. All T_{CH} values observed in the silica-bound AAP and AEP functionalities are smaller than 50 μs and only slightly increase toward their amide ends. These results show both these groups are immobilized on the NMR time scale, which in this case is given by the inverse of the ^{13}C-^1H dipolar coupling (i.e. $\sim(20kHz)^{-1}$).[35] In the CP-MSN sample, the longer T_{CH} values of 66, 119 and 2280 μs were observed for C1, C2-3 and C4, respectively, suggesting that the cyanopropyl functionalities experience increased mobility toward the nitrile end, which weakens the ^1H-^{13}C dipolar interactions and impedes the cross-polarization process. Due to the heterogeneity of the MSN surface, the exact nature of molecular mobility cannot be discerned based on these data. Note, however, that fast rotation about an axis parallel to the Si–C1 bond would lower the polarization rate by a factor of approximately two with respect to fully rigid case,[36] which is roughly what we observed (Table 1). The nitrile end of the molecule undergoes a more 'isotropic' motion, as indicated by the further increased values of T_{CH} for C2 and C3. For carbon C4 in CP-MSN the polarization transfer is additionally inhibited due to the lack of directly attached proton.

2.4.2 Computational Results.

Model with one functional group. Figure 5 shows the optimized structures and the ReaxFF relative energies for the prone and upright conformations of each of the studied groups on 4L-MSN. The AAP and AEP functionalities in upright conformations have a much higher energy, by 53 kcal/mol and 58 kcal/mol, respectively, than those in the prone positions. In contrast, the prone and upright conformations for CP are separated only by 9 kcal/mol, with the prone structure lower in energy. For the current level of theory, the energy difference of 9 kcal/mol is not considered to be significant. In this context, these computational results are in excellent agreement with the previously shown results of solid-state ^{13}C NMR measurements, which suggested that both AAP and AEP groups are motionless on the NMR time scale, whereas the CP groups are quite mobile.

Model with two functional groups. Considering the molecular lengths and surface concentrations of the AAP and AEP functionalities, it is expected that two neighboring groups are likely to interact with each other by forming hydrogen bonds through their amine moieties. In order to assess the resulting conformational changes, ReaxFF calculations were performed for the model in which the two functional groups are placed on 4L-MSN in close proximity, facing each other.

Figure 6 shows the optimized conformations and the corresponding relative energies for AAP-MSN models with two mutually interacting AAP groups at distances P-X, P-Y and P-Z, as described earlier (See Figure 2e). The calculations were performed using two AAP groups on sites P and X in the prone and upright conformations (Figure 2; Figures 6A and 6B). The prone (A) conformation has the lowest energy. The energies of the B, C, D and E structures are higher by 88, 78, 52 and 85 kcal/mol, respectively. These very high relative

energies mean that these structures are unlikely to exist. This suggests that the conformation of AAP is predominantly determined by the interaction between the functional groups and the silica surface, in agreement with the low mobility observed in the NMR experiment. The differences among C, D and E suggest that the AAP-AAP interactions depend strongly on the distances between the bonding sites, with the lowest energy structure (D) corresponding to the intermediate P-Y distance. In models C and D, the terminal amine of one AAP group is located close to another AAP molecule, suggesting the formation of a hydrogen bond. Especially in model D, the terminal amine of one AAP is very close to the dialkylamine of another AAP molecule. This conformation is suitable to form strong hydrogen bonds, which is most likely responsible for lowering the energy.

Figure 7 shows the optimized conformations and the corresponding relative energies for the AEP-4L-MSN models. The energies of conformations B, C, D and E are higher than that of the prone species A by 100, 115, 116, 70 and kcal/mol, respectively. The AEP molecules prefer the prone position, similar to the AAP-MSN. Unlike the AAP-MSN, two of the interacting conformations, i.e., C and D, have higher energy than the upright one. In contrast, the energy of conformation E is lower by 30 kcal/mol than that of B, likely due to the hydrogen bonding. Indeed, the terminal amine of one AEP molecule is located close to the dialkylamine of another AEP.

In both AAP-MSN and AEP-MSN, the conformations of functional groups are determined by the interactions between the amine moieties and the surface silanols. This result also suggests the possibility of controlling the conformations of functional groups by changing the properties of the surface, for example by modifying the number of silanols or

by replacing silanol protons with other cations. In other words, MSNs can play an active role as a co-catalyst.

2.5 Conclusions

The ReaxFF and ^{13}C NMR results provide valuable insights into the inner pore environment of MSNs. The ReaxFF method predicts that the AAP and AEP functional groups are primarily in the prone conformation on the silica surface and their motions are restricted by hydrogen bonds between the molecules and silanol groups. On the other hand, the CP groups show little preference between the prone and upright conformations, which suggests that they can rapidly interchange between these two conformations unless there is a large energy barrier between them. The solid-state ^{13}C NMR experiments show the same trends, suggesting that the AAP and AEP groups are static on the NMR time scale, whereas the CP groups exhibit considerable rotational motion about an axis parallel to the Si–C1 bond.

Both the active sites and the substrate surface play an important role in heterogeneous catalysis. ReaxFF appears to be a viable computational method for obtaining chemical properties of the MSN materials, and provides a useful alternative to more accurate ab initio quantum mechanics methods that are too computationally costly. The existence of local minima on each potential energy surface that has been explored in this work implies that there are energy barriers separating these species, in each case, from the global minima. These energy barriers have not been determined. The ReaxFF method will enable the study of more complex MSN systems with hundreds to thousands of atoms at a very low computational cost.

2.7 References.

[1]Beck, J. S.; Vartuli, J. C.; Roth, W. J.; Leonowicz, M. E.; Kresge, C. T.; Schmitt, K. D.; Chu, C. T. W.; Olson, D. H.; Sheppard, E. W.; McCullen, S. B.; Higgins, J. B.; Schlenker, J. L. *J. Am. Chem. Soc.* **1992,** *114*, 10834-10843.

[2]Sayari, A. *Chem. Mater.* **1996,** *8,* 1840-1852.

[3]Corma, A. *Chem. Rev.* **1997,** *97*, 2373-2419.

[4]Hoffmann, F.; Cornelius, M.; Morell, J.; Froba, M. *Angew. Chem. Int. Ed.* **2006,** *45*, 3216-3251.

[5]Ying, J. Y.; Mehnert, C. P.; Wong, M. S. *Angew. Chem. Int. Ed.* **1999,** *38*, 56-77.

[6]Ruiz-Hitzky, E.; Aranda, P.; Darder, M.; Ogawa, M. *Chem. Soc. Rev.* **2011,** *40*, 801-828.

[7]Mao, K.; Kobayashi, T.; Wiench, J. W.; Chen, H.-T.; Tsai, C.-H.; Lin, V. S.-Y.; Pruski, M. *J. Am. Chem. Soc.* **2010,** *132*, 12452-12457.

[8]Møller, C.; Plesset, M. S. *Phys. Rev.* **1934,** *46*, 618-622

[9]Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964,** *136*, B864-B871.

[10]Zhang, R. Q.; Chu, T. S.; Lee, S. T. *J. Chem. Phys.* **2001,** *114*, 5531-5536.

[11]Civalleri, B.; Casassa, S.; Garrone, E.; Pisani, C.; Ugliengo, P. *J. Phys. Chem. B* **1999,** *103*, 2165-2171.

[12]Sauer, J.; Hill, J. R. *J. Phys. Chem.* **1994,** *98*, 1238-1244.

[13]Netzloff, H. M.; Collins, M. A. *J. Chem. Phys.* **2007,** *127*, 134113.

[14]Liu, B.; Wang, W.; Zhang, X. *Phys. Chem. Chem. Phys.* **2004,** *6*, 3985–3990.

[15]Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989,** *111*, 8551-8566.

[16]Allinger, N. L.; Li, F.; Yan, L. *J. Comput. Chem.* **1990,** *11*, 848-867.

[17]Allinger, N. L.; Li, F.; Yan, L.; Tai, J. C. *J. Comput. Chem.* **1990,** *11*, 868-895.

[18]Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990,** *94*, 8897-8909.

[19]van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. *J. Phys. Chem. A* **2001,** *105*, 9396-9409.

[20]van Duin, A. C. T.; Strachan, A.; Stewman, S.; Zhang, Q.; Xu, X.; Goddard, W. A. *J. Phys. Chem. A* **2003,** *107*, 3803-3811.

[21]Huh, S.; Chen, H.-T.; W., W. J.; M., P.; Lin, S.-Y. V. *J. Am. Chem. Soc.* **2004,** *126,* 1010-1011.

[22]Huh, S.; Wiench, J. W.; Yoo, J. C.; Pruski, M.; Lin, V. S. Y. *Chem. Mater.* **2003,** *15*, 4247-4256.

[23]Trebosc, J.; Wiench, J. W.; Huh, S.; Lin, V. S. Y.; Pruski, M. *J. Am. Chem. Soc.* **2005,** *127*, 7587-7593.

[24]Wiench, J. W.; Lin, V. S. Y.; Pruski, M. *J. Magn. Reson.* **2008,** *193*, 233-242.

[25]Bennett, A. E.; Rienstra, C. M.; Auger, M.; Griffin, R. G. *J. Chem. Phys.* **1995,** *103*, 6951-6958.

[26]Mehring, M. *Principles of High Resolution NMR in Solids*. 2nd ed.; Springer-Verlag: New York, 1983.

[27]Chenoweth, K.; Cheung, S.; T., v. D. A. C.; A., G. W.; Kober, E. M. *J. Am. Chem. Soc.* **2005,** *127*, 7192-7202.

[28]Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S. J., J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993,** *14*, 1347-1363.

[29]Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C. E.; Frenking, G.; Kim, K. S.; Scuseria, G. E., Eds.; Elsevier, Amsterdam, 2005, p. 1167-1189.

[30]Peacor, D. R. *Z. Kristallogr.* **1973,** *138*, 274-298.

[31]Frondel, C., *Am. Mineral.* **1979,** *64*, 799-804.

[32]Kobayashi, T.; DiVerdi, J. A.; Maciel, G. E. *J. Phys. Chem. C* **2008,** *110*, 4315-4326.

[33]Maciel, G. E. In *Encyclopedia of Nuclear Magnetic Resonance*, Grand, M. D.; Harris, R. K., Eds.; John Wiley & Sons Ltd.: Chichester, U.K., 1996; p. 4370.

[34]http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre_index.cgi

[35]Amoureux, J. P.; Pruski, M. *Mol. Phys.* **2002,** *100*, 1595-1613.

[36]Slichter, C. P. *Principles of Magnetic Resonance*. 2nd ed.; Springer-Verlag: New York, 1980.

Figure 1. Diagram of functional groups, AAP, AEP and CP. The carbon atom numbering system is also shown. Carbon C1 is covalently bonded to a silicon atom on the silica surface.

CP   AAP   AEP
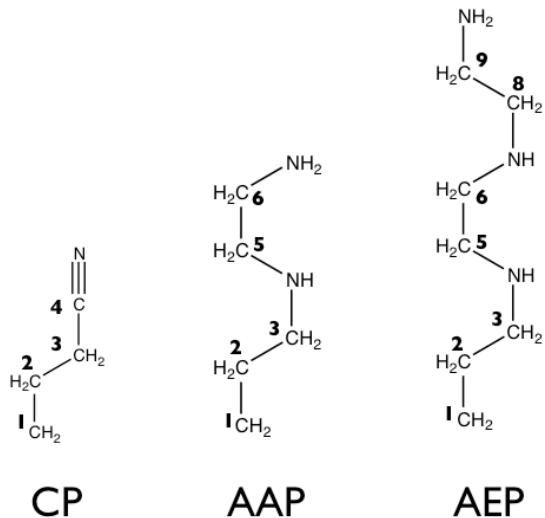
Figure 2. Model surface of inner pore of MSNs. (a) TEM image of an MSN. (b) Expanded section of a TEM image showing side view of the pore channels. (c) Schematic side view of MSN pore showing attached catalytic substituents (in blue) to the inner pore surface. (d) Side view of the four-layer silica slab (4L-MSN). (e) Top view of the model surface, where the site P holds one of the functional groups and the site X, Y or Z holds a neighboring group. The square surrounding P represents a 1 nm^2 surface area.
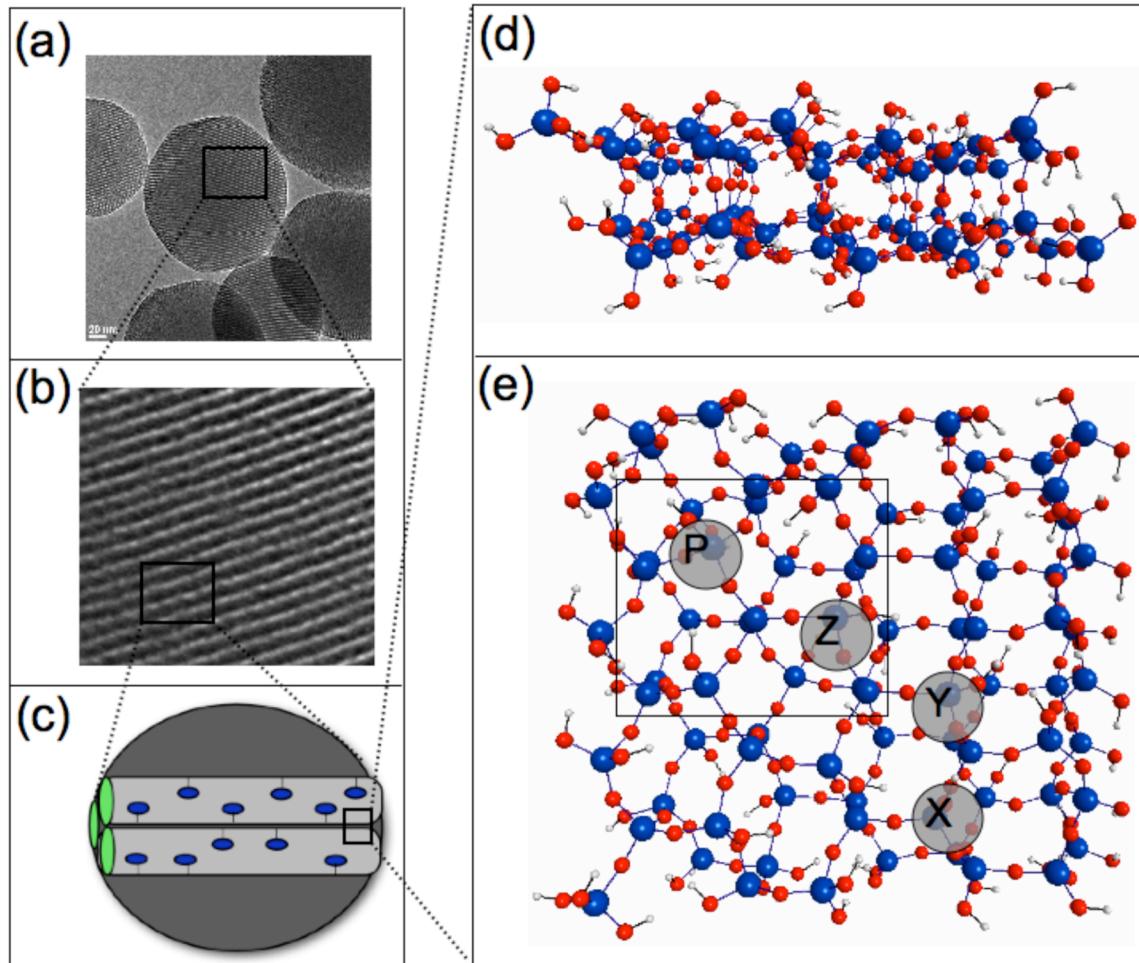
Figure 3. ^{29}Si DPMAS spectra of (a) CP-MSN, (b) AAP-MSN and (c) AEP-MSN.
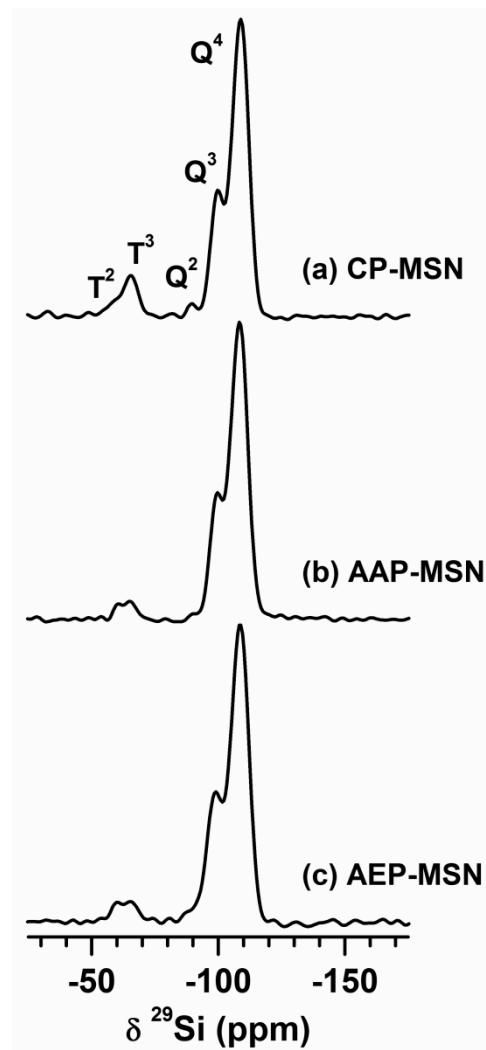
Figure 4. $^{13}C\{^1H\}$ CPMAS spectra of (a) CP-MSN, (b) AAP-MSN and (c) AEP-MSN, measured using τ_{CP} = 100 µs. Asterisk in the spectrum of CP-MSN represents the residual surfactant carbons.
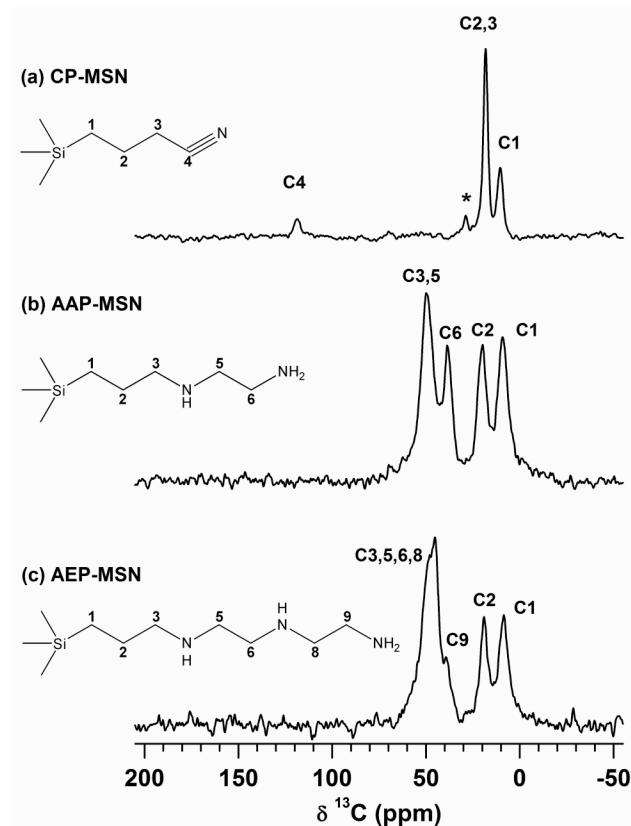
Figure 5. ReaxFF optimized conformations and corresponding relative energies of one AAP (A, B), AEP (C, D), and CP (E, F) molecule on 4L-MSN, in upright (A, C, E) and prone (B, D, F) conformations. The prone substituent lies along the 4L-MSN surface and orients towards the viewer. Energies are in kcal/mol and are referred to the lower-energy conformations. "P" refers to the position P in Figure 2e.
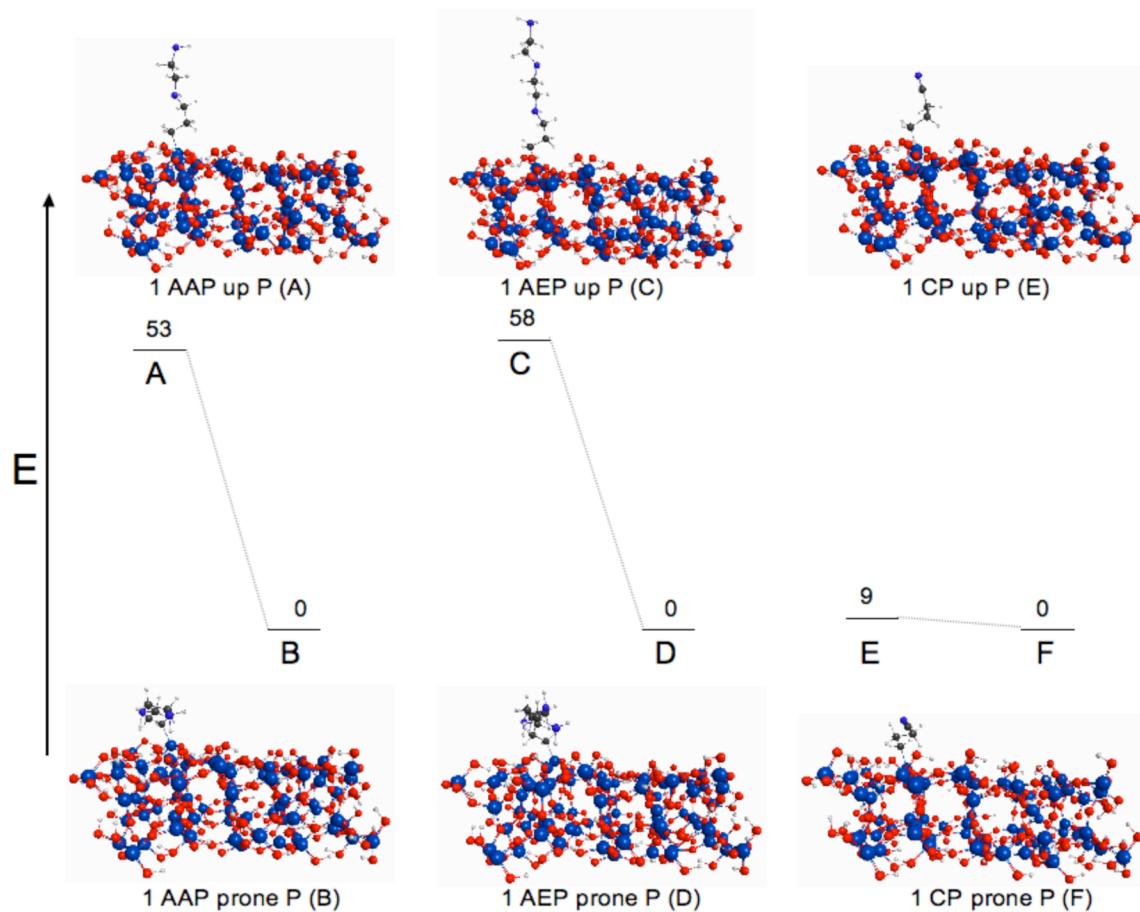
Figure 6. ReaxFF optimized conformations and relative energies of AAP on 4L-MSN. (A, B) show prone and upright conformations, respectively, with two AAP molecules attached to the sites P and X. (C, D, E) represent three interacting conformations in which the site P holds one AAP and the site X, Y or Z holds another, respectively. The substituent attached to position P orients towards the viewer above the 4L-MSN surface, while the substituent attached to sites X, Y, or Z orients away from the viewer above the 4L-MSN surface. Energies are in kcal/mol relative to the lowest energy conformation (A).

2 AAP up P-X (B)

2 AAP hbond P-Z (E)

2 AAP hbond P-X (C)

2 AAP hbond P-Y (D)

2 AAP prone P-X (A)
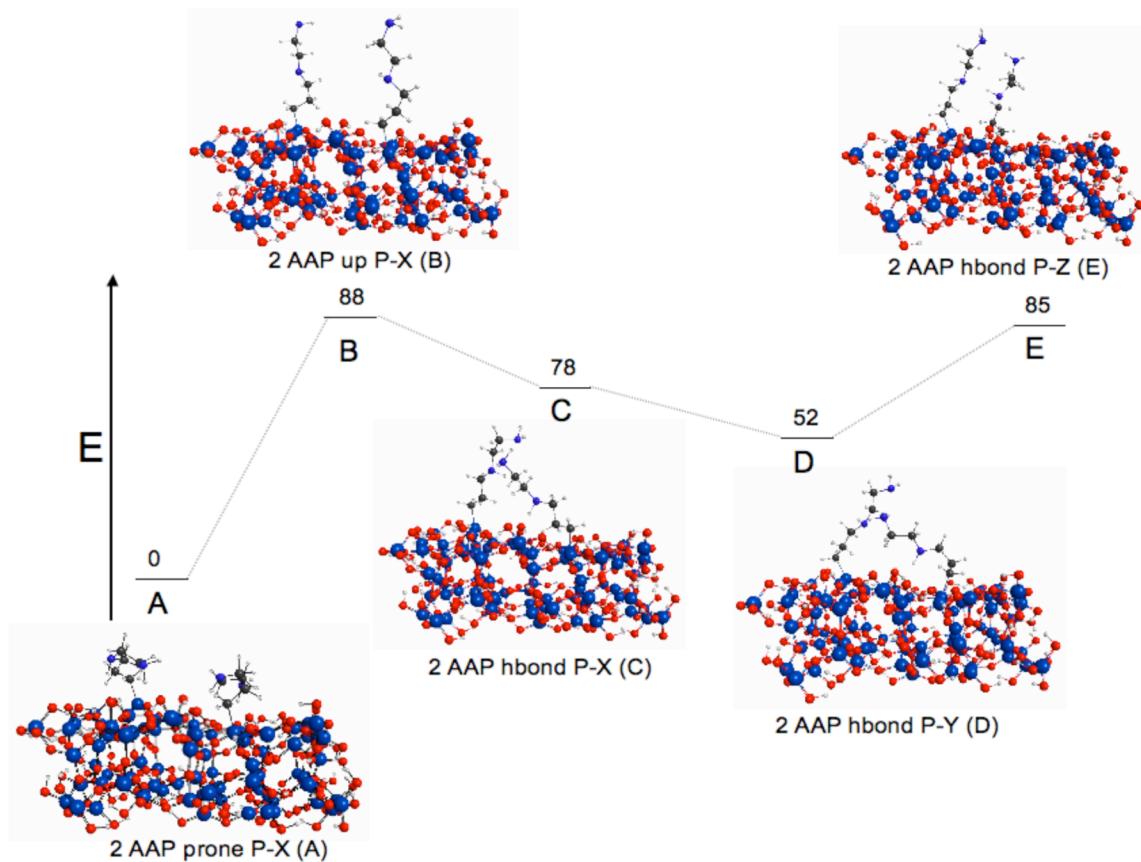
Figure 7. ReaxFF optimized conformations and relative energies of pairs of AEP functional groups on 4L-MSN. (A, B) show prone and upright conformations, respectively, with two AEP molecules attached to sites P and X. (C, D, E) represent three interacting conformations in which the site P holds one AEP and the site X, Y or Z holds another, respectively. The substituent attached to position P orients towards the viewer above the 4L-MSN surface, while the substituent attached to sites X, Y, or Z orients away from the viewer above the 4L-MSN surface. Energies are in kcal/mol relative to conformation (A).
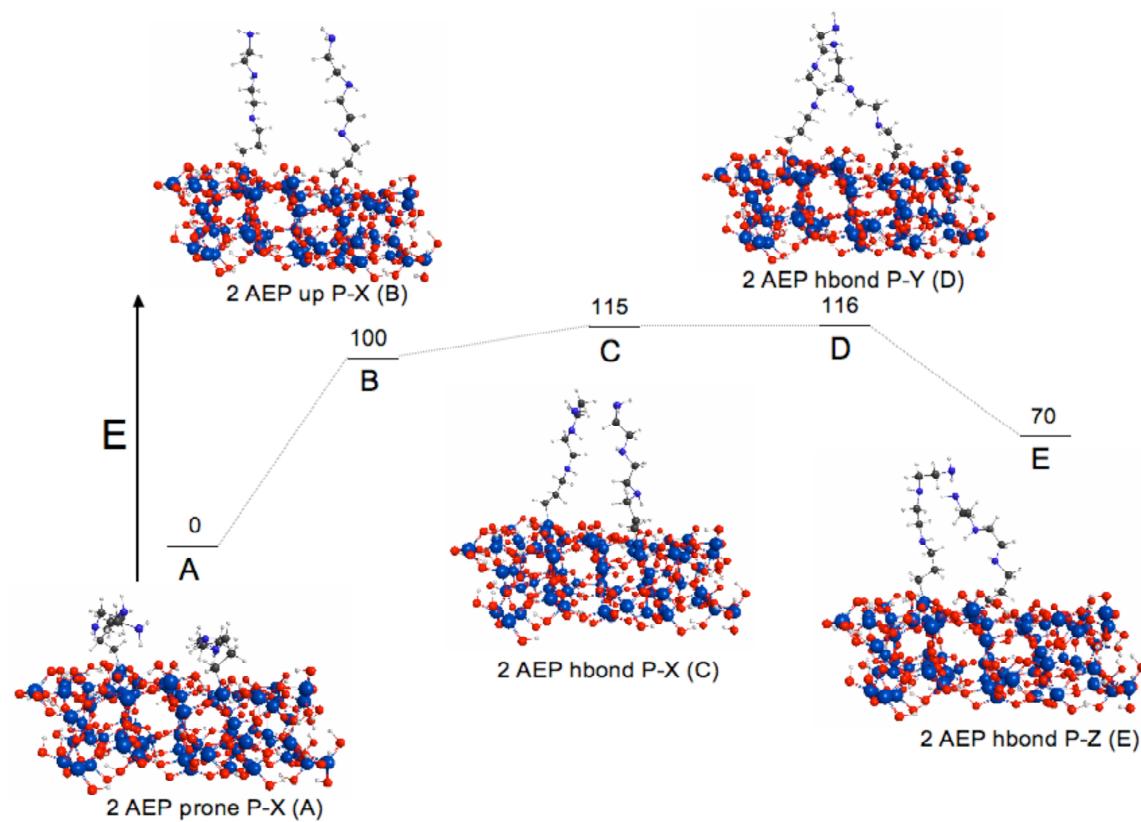
Table 1. Relative concentrations of T^n and Q^n sites, and the amounts of organic functionalities estimated from ^{29}Si DPMAS NMR spectra.

Sample	relative concentrations				amount of functionalities	
	T^2+T^3	Q^2	Q^3	Q^4	(in mmol/g)	(in $1/nm^2$)[a]
CP-MSN	11	1	23	66	1.4	0.8
AAP-MSN	6	1	26	67	0.9	0.8
AEP-MSN	4	1	23	72	0.7	0.5

[a] Estimated using the following BET surface areas; CP-MSN: 1013 m^2/g, AAP-MSN: 665 m^2/g, AEP-MSN: 806 m^2/g.[21,22]

Table 2. Cross polarization times T_{CH} (μs) for individual carbons in the CP-, AAP- and AEP-MSN samples.

CP-MSN		AAP-MSN		AEP-MSN	
C1	66	C1	25	C1	26
C2,3	119	C2	29	C2	35
C4	2280	C3,5	28	C3,5,6,8	37
		C6	33	C9	48

Chapter 3

SIMOMM-RX: A QM/MM METHOD WITH A REACTIVE FORCE FIELD

A paper to be submitted to the Journal of Chemical Physics

Sean A. Nedd and Mark S. Gordon

3.1 Abstract

The Reactive Force Field (ReaxFF) has been incorporated into the Surface Integrated Molecular Orbital/Molecular Mechanics (SIMOMM) method. The new method called SIMOMM-Rx is compared to other hybrid Quantum Mechanical/Molecular Mechanical (QM/MM) methods including the Weiner method and the original SIMOMM method. SIMOMM-Rx optimizations were done on the silicon clusters SI_9H_{12} and $Si_{38}H_{36}$, both of which have a Si_6H_{12} QM region. For these silicon clusters, SIMOMM-Rx predicts bond lengths (± 0.03 Å), bond angles (±6.0°), and torsion angles (±9.0°) that are similar to those predicted by the other QM/MM methods. The SIMOMM-Rx method, with density functional theory and second order perturbation theory employed as the QM part, is used to predict the structures of a single catalytic substituent on a four-layer silica slab that represents the surface of the inner pore of a mesoporous silica nanosphere (MSN), in order to assess the lowest energy conformer of the substituent on the slab surface. The catalytic substituents are N-(2-aminoethyl)-3-aminopropyl- (AAP), N-[N-(2-aminoethyl)-2-aminoethyl]-3-aminopropyl- (AEP) and 3-cyanopropyl- (CP). In agreement with previous calculations, the SIMOMM-Rx calculations predict that the AAP and AEP groups exist mainly in the prone positions while the CP group can exist in the prone or upright position.

3.2 Introduction.

Mesoporous silica nanospheres (MSNs), such as MCM-41, have shown promising potential for applications in heterogeneous catalysis. This is due to their large surface area, well defined pore structure and narrow pore size distribution. [1-3] The experimental production and instrumental analysis of the resulting MSN materials, in which catalytic substituents are immobilized within the inner pore surfaces, have been studied and reviewed. [3-5] However, there is little understanding of the structure and behavior of catalytic functionalities within these pores due to the amorphous nature of these silica systems, which are difficult to characterize using typical experimental techniques. Thus, computational methods are useful in understanding such amorphous system.

In a previous computational study [6], three organic functional groups, N-(2-aminoethyl)-3-aminopropyl- (AAP), N-[N-(2-aminoethyl)-2-aminoethyl]-3-aminopropyl- (AEP) and 3-cyanopropyl- (CP) (Figure 1), were covalently attached to an inner pore surface model of the MSNs, and their conformations were studied using the ReaxFF force field [7,8] theoretical calculations and then validated by comparison with solid-state ^{13}C NMR experiments.

The bond-order dependent ReaxFF force field was used because of its ability to treat silica systems with hundreds to thousands of atoms at low computational cost. [6] The MSN model used was a four-layer MSN silica slab (4L-MSN). Attaching one of these functional groups (AEP, AAP, and CP) to 4L-MSN gives three substituted MSN systems: AAP-MSN, AEP-MSN, and CP-MSN respectively. The lowest energy MSN system was determined by comparing the functional group in an upright position versus a prone (parallel across the surface of the 4L-MSN) position on the 4L-MSN surface. This ReaxFF based study indicated

that the prone structure is favored for the AAP-MSN, and AEP-MSN models, and that there is little difference energetically between the two conformers for CP-MSN. These ReaxFF results were correlated with solid-state ^{13}C NMR mobility data for AAP-MSN, AEP-MSN, and CP-MSN systems. The mobility data refers to the intensity of excitation of the last carbon on the ends of the organic functional groups or the terminal carbon as shown in Figure 1. Focusing on the mobility data, AAP-MSN and AEP-MSN showed low mobility, while CP-MSN showed two orders of magnitude higher mobility. The low mobility AAP-MSN and AEP-MSN ^{13}C NMR data were correlated to the prone conformer having a lower energy than the upright conformer, however, the high mobility CP-MSN data was correlated to a small difference in energies between the prone and upright conformers. [6]

Since the ReaxFF based structures were correlated with mobility data, ReaxFF was successfully used to gain insight into the structure of catalytic substituents on the MSN inner pore surface. However, even though ReaxFF offers reasonable relative energies, ReaxFF is not optimized to take into consideration all types of interactions within a silica system. For example, ReaxFF may not properly treat the weaker interactions prevalent in MSN systems due to the many hydroxyl groups on the silica surface, such as hydrogen bonding and dispersion forces. Since ReaxFF is parametrized using density functional theory with the B3LYP functional[9], DFT/B3LYP (and thus ReaxFF) is not sufficiently accurate for such weak interactions involving dispersion forces. [10,11] In addition, the benchmark structures used to develop ReaxFF the ReaxFF parameters for atom pairs that are relevant to MSN species did not include weaker interactions.

In order to systematically improve on the investigation of structures on the surface of the inner pore of MSN, higher-level (more accurate) computational methods are desirable.

However, these higher-level computational methods generally incur a high computational cost (in terms of computational time and resources). One solution is to employ the hybrid QM/MM (quantum mechanics/molecular mechanics) technique. These QM/MM techniques use a QM method for the active region within a chemical system, while the remainder of the system uses a MM method. The specific portion of interest is termed the reactive site model (RSM) and the remainder of the system is called the bulk model (BM). QM/MM methods that are relevant to surface science include the Weiner[12,13] method and SIMOMM (surface integrated molecular orbital/molecular mechanics). [14] Further insight into MSN catalysis can be gained with a QM/MM method like SIMOMM, which was designed to treat the surfaces of chemical systems. The SIMOMM method is appealing, because it minimizes edge effects by expanding the size of a system of interest, at minimal computational cost and with reasonable accuracy [14-24]. The SIMOMM technique was originally implemented as an interface between the GAMESS (General Atomic and Molecular Electronic Structure System) [25] for the QM part and Tinker [26-28] for the MM part of the QM/MM calculations. The present work describes a new version of SIMOMM, called SIMOMM-Rx, that employs the ReaxFF reactive force field for the MM part. The ReaxFF force field has previously been implemented into GAMESS and employed to study the structures of substituted MSN model species. [6]

The viability of the SIMOMM-Rx method is first assessed by comparing the new method with the original SIMOMM method [14] for small silicon clusters (Si_9H_{12} and $Si_{38}H_{36}$) that have been used to represent the Si(100) surface [16-19,21,23,24,29]. Comparisons are also presented with the method of Weiner and co-workers [12]. In addition, the SIMOMM-Rx method is used to analyze the AEP-MSN, AAP-MSN, and CP-MSN structures. [6]

3.3 Computational Details.

3.3.1 SIMOMM-Rx. The determination of the SIMOMM-Rx optimized structures require the gradients of the energy in three regions of the QM/MM system: the RSM (Region 1), the BM (Region 2), and the capping hydrogen atoms that are added to the RSM when the BM is not present (Region 3). Examples of RSM and BM regions for silicon clusters are shown in Figure 2. The coordinates for Regions 1, 2, and 3 are referred to as R_1, R_2, and R_3, respectively. R_1 and R_3 are both optimized using both QM and MM, while R_2 is optimized using only MM. The optimization is achieved via iterations over a hybrid QM/MM gradient involving the QM and MM energies (E_{QM} and E_{MM}, respectively), and the related coordinates in R_1, R_2, and R_3. The total energy (E_T) is

$$E_T = E_{QM} + E_{MM} . \tag{1}$$

The SIMOMM hybrid gradient is split into three parts based on the three regions [14]:

$$\frac{dE_T}{dR_1} = \frac{\partial E_{QM}}{dR_1} + \frac{\partial E_{MM}}{dR_1} , \tag{2}$$

$$\frac{dE_T}{dR_3} = \frac{\partial E_{QM}}{dR_3} , \tag{3}$$

$$\frac{dE_T}{dR_2} = \frac{\partial E_{MM}}{dR_2} . \tag{4}$$

In order to obtain the SIMOMM-Rx energy, the following procedure is followed [14]:

1. The chemical structure is separated into Region 1 atoms, and Region 2 atoms. Then the capping H atoms from Region 3 are added to Region 1 to give a new Region 4.

2. A QM energy optimization is initiated on the Region 4 atoms. The QM gradients for Region 1 only are saved. The energy for the Region 4 atoms is saved as the QM energy.

3. The optimized QM coordinates (Region 1 only) are added to the Region 2 coordinates. This combined system is then optimized using ReaxFF; the Region 1 coordinates are frozen during the ReaxFF optimization. The MM gradients for Regions 1 and 2 are saved.

4. A ReaxFF single point energy calculation is done on the Region 2 atoms. The resulting energy is saved as the MM energy.

5. The final QM/MM energy is the combination of the QM and MM energy values.

6. The hybrid QM/MM gradient is the summation of the saved gradients from Regions 1 and 2. The QM/MM gradient is compared against a convergence criterion (default = 10^{-4} hartree/bohr for QM and 0.001 kcal/mol/angstrom for MM) that can be defined by the user. If the criteria are not satisfied, steps 1-6 are repeated.

All ReaxFF computations were done using the ReaxFF$_{SiO/Si}$ version. [8,30] The ReaxFF optimizations use the conjugate gradient converging scheme, while the QM methods use analytic gradients.

Silicon Clusters. In order to test the SIMOMM-Rx method, two silicon clusters, each with a single surface dimer, were used. Structural optimizations using SIMOMM-Rx were run on the Si_9H_{12} and the $Si_{38}H_{36}$ cluster shown in Figure 2. In both clusters, Si_6H_{12} (RSM) was calculated using the General Valence Bond with perfect Pairing (GVB-PP(1)) [31,32] theory with the Hay-Wadt Effective Core Potential (HW ECP) [33] basis set, while the remainder of each cluster (the BM) was treated with ReaxFF. The geometric parameters given in Tables 1 and 2 use the atom numbering system in Figure 2.

Mesoporous silica nanospheres. SIMOMM-Rx was used to characterize the behavior of organic functional groups on the surface of the inner pore of MSNs. The QM methods used

for these MSN calculations employ the methods of DFT/B3LYP [9], DFT/PBE0 [34], and second order perturbation theory (MP2) [35] all using the 6-311+G(d) basis set[36]. Single point energy SIMOMM-Rx calculations were performed on models consisting of a four-layer silica slab, referred to as 4L-MSN (Figures 3a and 3b), and an attached organic functional group (AAP, AEP, or CP, Figure 1). The 4L-MSN species represent the inner surface of functionalized MSNs (AAP-MSN, AEP-MSN, and CP-MSN). The MSN starting structures were obtained from the ReaxFF structures determined previously. [6,37,38] A single functional group was attached to a site labeled "P" in Figure 3b, using two conformations that represent the prone (parallel) and perpendicular (upright) orientations relative to the 4L-MSN surface. Since the prone and upright conformer energy effects need to be isolated, the same 4L-MSN backbone structure [6] is used for all MSN systems (AAP-MSN, AEP-MSN, and CP-MSN). In the SIMOMM-Rx calculations, the RSM includes the functional group and part of the 4L-MSN as indicated in Figure 3c. Figure 3c illustrates the example in which -Si(OSi)$_3$ from the 4L-MSN is attached to the AEP functional group. This -Si(OSi)$_3$ portion of the 4L-MSN is also used as part of the RSM for the AAP-MSN and CP-MSN systems. The SIMOMM-Rx structures obtained in this manner were used to determine the relative stabilities for the three catalytic substituents in the upright vs. the prone orientation and compared to the ReaxFF results [6].

3.4 Results and Discussion.

Tables 1 and 2, columns 2, 3, and 4 represent QM/MM calculations for the Weiner, SIMOMM/Tinker, and SIMOMM-Rx respectively. Columns 5 and 6 represent the QM based treatment of all atoms using the GVB-PP(1) method on only the Si$_6$H$_{12}$ structure, and the

whole cluster respectively. Columns 7 and 8 represent the MM based treatment of all atoms using Tinker and ReaxFF, respectively. The RSM (QM) row indicates the atoms used for the RSM, while the BM (MM) row indicates the whole cluster. The BM contains those atoms that are not in the RSM region. All data besides columns 4 and 8 were obtained from Ref. 14. The bond length of the dimer on the surface of the Si_9H_{12} and $Si_{38}H_{36}$ silicon clusters (2-1 in Tables 1 and 2) is important for assessing the accuracy of these theoretical methods. It has been previously shown [14] that within these silicon clusters, the Tinker dimer bond lengths of 2.376 Å in Table 1 and 2.370 Å in Table 2, are closer to a single bond length than to a double bond length. [14] The GVB-PP(1) method predicts a more accurate smaller dimer bond length of 2.249 Å. Using the GVB-PP(1) dimer bond length as a reference, the QM/MM methods of Weiner, SIMOMM/Tinker, and SIMOMM-Rx all predict 2-1 bond lengths within 0.06 Å. These calculated bond lengths suggest that the SIMOMM based calculations correctly capture the dimer bonding.

Figure 4 shows the ReaxFF optimized structures from Ref. 6 for the prone and upright conformations of each of the studied groups on 4L-MSN. Figure 4 also shows the relative energies of the prone and up conformations obtained using ReaxFF, as well as single point SIMOMM-Rx energies. SIMOMM-Rx with all three QM methods predicts that the prone conformation is lower in energy than the upright conformation, in agreement with the previous ReaxFF calculations. However, the SIMOMM-Rx relative energies are much smaller than the ReaxFF relative energies for the AAP and AEP substituents. For CP-MSN, there is no hydrogen bonding with the 4L-MSN surface to consider, unlike the cases of AAP-MSN and AEP-MSN; this lack of hydrogen bonding in CP-MSN is reflected in the SIMOMM-Rx relative energies, which only differ by 4 kcal/mol from the ReaxFF relative

energies, as shown in Figure 4. The larger ReaxFF relative energies for AAP and AEP may be due to the ReaxFF parameter set, which is not fully optimized for systems that contain weak electrostatic interactions within the MSN structures. The ReaxFF method may consequently overbind the interactions between the silica surface and the catalytic substituent when the substituent is in the prone position. According to SIMOMM-Rx, AEP-MSN, AAP-MSN, and CP-MSN all favor the prone position, with the CP preference slightly smaller than it is for AAP. The AEP substituent most strongly favors the prone position, by 10 kcal/mol. It is likely that SIMOMM-Rx represents a more accurate representation of the magnitude of relative energies than does ReaxFF by itself. It is noteworthy that the three QM methods, DFT/B3LYP, DFT/PBE0, and MP2 are all in very good agreement with each other with regard to the relative energies of the prone vs. up conformations.

3.5 Conclusions

The SIMOMM-Rx method has been tested on the Si_9H_{12} and $Si_{38}H_{36}$ clusters against other QM/MM techniques (Weiner and SIMOMM-Tinker), as well as pure quantum mechanical (GVB-PP(1)) and molecular mechanics methods (Tinker). The dimer bond lengths in these silicon clusters only differ by ±0.03 Å between SIMOMM-Rx and the other QM/MM methods. The multireference nature of the silicon cluster systems makes predicting consistent results difficult, and SIMOMM-Rx, with the GVB-PP(1) QM method successfully reproduces results that were published previously.

SIMOMM-Rx predicts the ReaxFF trends for the prone vs. up relative energies, although the variation in these relative energies as a function of substituent (AAP vs. AEP vs. CP) are considerably attenuated in comparison with the previously reported ReaxFF

calculations. The calculated SIMOMM-Rx relative energies are nearly independent of the QM method used. DFT/B3LYP, DFT/PBE0, and MP2, all with the 6-311+G(d) basis set, predict very similar prone vs. up relative energies. Of the three substituents, CP results in the smallest prone vs up relative energy, in agreement with ReaxFF, but the variation with substituents is small.

SIMOMM-Rx provides a useful tool for the study of systems like silica nanoporous materials, for which fully QM calculations are computationally challenging. Further extensions of SIMOMM-Rx will include adding a method that will enable the study of solvent effects on surface science.

3.7 References.

[1]J. S. Beck, J. C. Vartuli, W. J. Roth, M. E. Leonowicz, C. T. Kresge, K. D. Schmitt, C. T. W. Chu, D. H. Olson, E. W. Sheppard, S. B. McCullen, J. B. Higgins, and J. L. Schlenker, J. Am. Chem. Soc. **114**, 10834 (1992).

[2]A. Sayari, Chem. Mater. **8**, 1840 (1996).

[3]A. Corma, Chem. Rev. **97**, 2373 (1997).

[4]F. Hoffmann, M. Cornelius, J. Morell, and M. Froba, Angew. Chem. Int. Ed. **45**, 3216 (2006).

[5]J. Y. Ying, C. P. Mehnert, and M. S. Wong, Angew. Chem. Int. Ed. **38**, 56 (1999).

[6]S. Nedd, T. Kobayashi, T. Chih-Hsiang, I. I. Slowing, M. Pruski, and M. S. Gordon, J. Phys. Chem. C **115**, 16333 (2011).

[7]A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, J. Phys. Chem. A **105**, 9396 (2001).

[8]A. C. T. van Duin, A. Strachan, S. Stewman, Q. Zhang, X. Xu, and W. A. Goddard, J. Phys. Chem. A **107**, 3803 (2003).

[9]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[10]E. J. Meijer and M. Sprik, J. Chem. Phys. **105**, 8684 (1996).

[11]S. M. Cybulski and C. E. Seversen, J. Chem. Phys. **122**, 014117 (2005).

[12]B. Weiner, C. S. Carmer, and M. Frenklach, Phys. Rev. B **43**, 1678 (1991).

[13]C. S. Carmer, B. Weiner, and M. Frenklach, J. Chem. Phys. **99**, 1356 (1993).

[14]J. R. Shoemaker, L. W. Burggraf, and M. S. Gordon, J. Phys. Chem. **103**, 3245 (1999).

[15]P. Arora, W. Li, P. Piecuch, J. W. Evans, M. Albao, and M. S. Gordon, J. Phys. Chem. C **114**, 12649 (2010).

[16]M. K. Ghosh, M. I. M. Sarker, and C. H. Choi, J. Phys. Chem. C **112**, 9327 (2008).

[17]M. K. Ghosh and C. H. Choi, Chem. Phys. Lett. **426**, 365 (2006).

[18]M. K. Ghosh and C. H. Choi, J. Phys. Chem. B **110**, 11277 (2006).

[19]H. S. Lee, K.-S. An, Y. Kim, and C. H. Choi, J. Phys. Chem. B **109**, 10909 (2005).

[20]P. Zapol, L. A. Curtiss, H. Tamura, and M. S. Gordon, Computational Materials Chemistry, 266 (2004).

[21]C. Lim and C. H. Choi, J. Phys. Chem. B **107**, 6853 (2003).

[22]M. M. Hurley, J. B. Wright, G. H. Lushington, and W. E. White, Theoretical Chemistry Accounts **109**, 160 (2003).

[23]C. H. Choi and M. S. Gordon, J. Am. Chem. Soc. **124**, 6162 (2002).

[24]Y. Jung, C. H. Choi, and M. S. Gordon, J. Phys. Chem. B **105**, 4039 (2001).

[25]M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, and N. Matsunaga, Journal of Computational Chemistry **14**, 1347 (1993).

[26]J. W. Ponder and F. M. Richards, J. Comput. Chem. **8**, 1016 (1987).

[27]C. E. Kundrot, J. W. Ponder, and F. M. Richards, J. Comput. Chem. **12**, 402 (1991).

[28]P. Ren and J. W. Ponder, J. Phys. Chem. B **107**, 5933 (2003).

[29]H. S. Lee, K.-S. An, Y. Kim, and C. H. Choi, J. Phys. Chem. B **109**, 10909 (2005).

[30]K. Chenoweth, S. Cheung, v. D. A. C. T., G. W. A., and E. M. Kober, Journal of Chemical Society **127**, 7192 (2005).

[31]W. A. Goddard III and L. B. Harding, Ann. Rev. Phys. Chem. **29**, 363 (1978).

[32]W. R. Wadt and W. A. Goddard III, J. Am. Chem. Soc. **97**, 2034 (1975).

[33]P. J. Hay and W. R. Wadt, J. Chem. Phys. **82**, 270 (1985).

[34]M. Ernzerhof and G. E. Scuseria, J. Chem. Phys. **110**, 5029 (1999).

[35]C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).

[36]R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).

[37]D. R. Peacor, Z. Kristallogr **138**, 274 (1973).

[38]C. Frondel, Am. Mineral. **64**, 799 (1979).

Figure 1. The organic functional groups AAP, AEP and CP. The bottom most Carbon atom shown for each group is covalently bonded to a silicon atom on the silica surface. The topmost carbon shown for each group is the terminal carbon.

CP   AAP   AEP
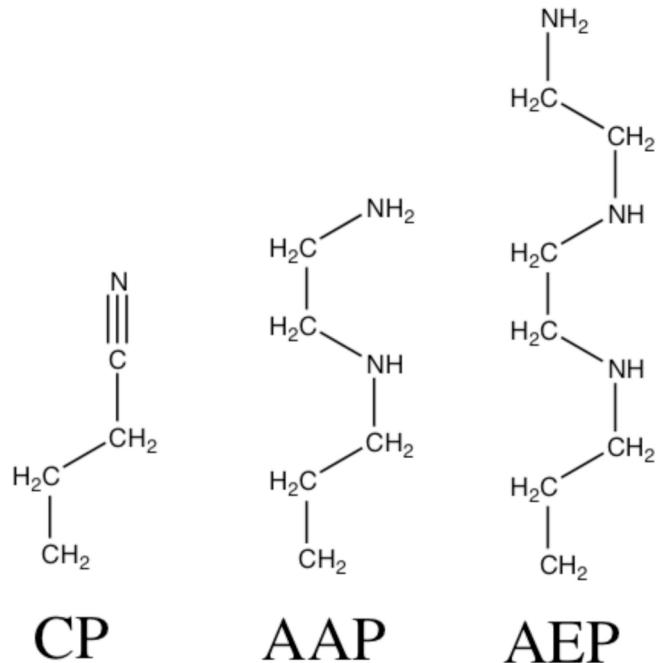
Figure 2. (a) Diagram of the Si_9H_{12} cluster. (b) Diagram of the $Si_{38}H_{36}$ cluster. The RSM and BM regions are highlighted in (a) and (b). The large spheres are Si atoms and the smaller spheres are H atoms.
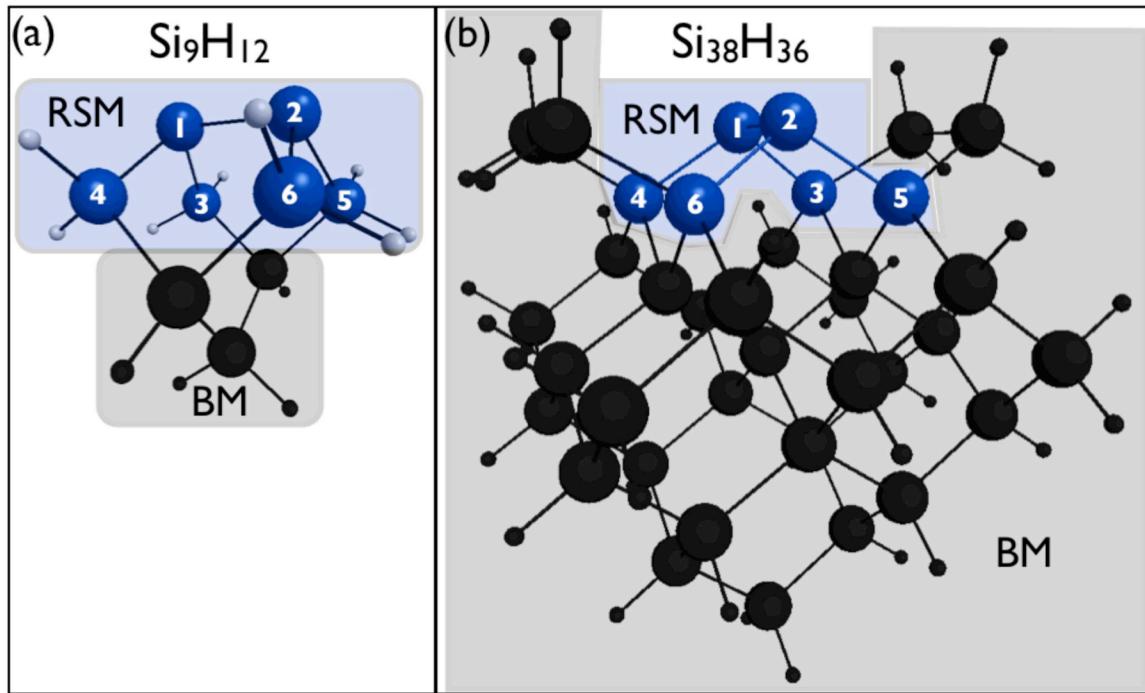
Figure 3. Model surface of inner pore of MSNs. (a) Side view of the four-layer silica slab (4L-MSN). (b) Top view of the model surface; the site P indicates the Si atom where a functional group is attached. The square surrounding "P" represents a 1 nm^2 surface area. (c) View of AEP-MSN showing the highlighted atoms representing the Reactive Site Model (RSM); the non-highlighted atoms represent the Bulk Model (BM). The region treated with a QM method is circled.
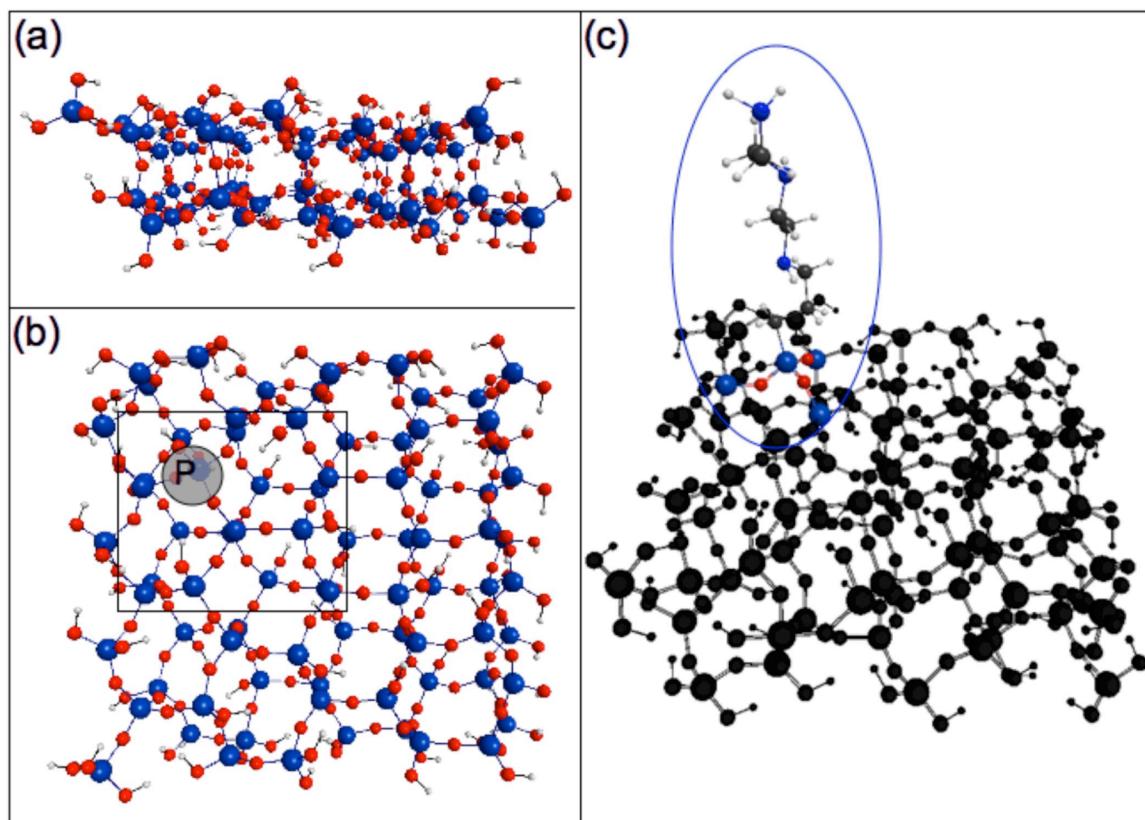
Figure 4. ReaxFF optimized conformations and the corresponding relative energies of one AAP (A, B), AEP (C, D), and CP (E, F) molecule on a 4L-MSN, in upright (A, C, E) and prone (B, D, F) conformations. SIMOMM-Rx relative energies using the QM methods (from top to bottom: B3LYP/6-311+G(d) (blue), PBE0/6-311+G(d) (green), and MP2/6-311+G(d) (red)) are indicated, with the ReaxFF relative energies in black. The regions treated with a QM method are circled. The prone substituent lies along the 4L-MSN surface and orients towards the viewer. Energies are in kcal/mol and are relative to the lower-energy prone conformations. The functional groups (AAP, AEP, and CP) are attached to 4L-MSN at the position P as shown in Figure 3b.
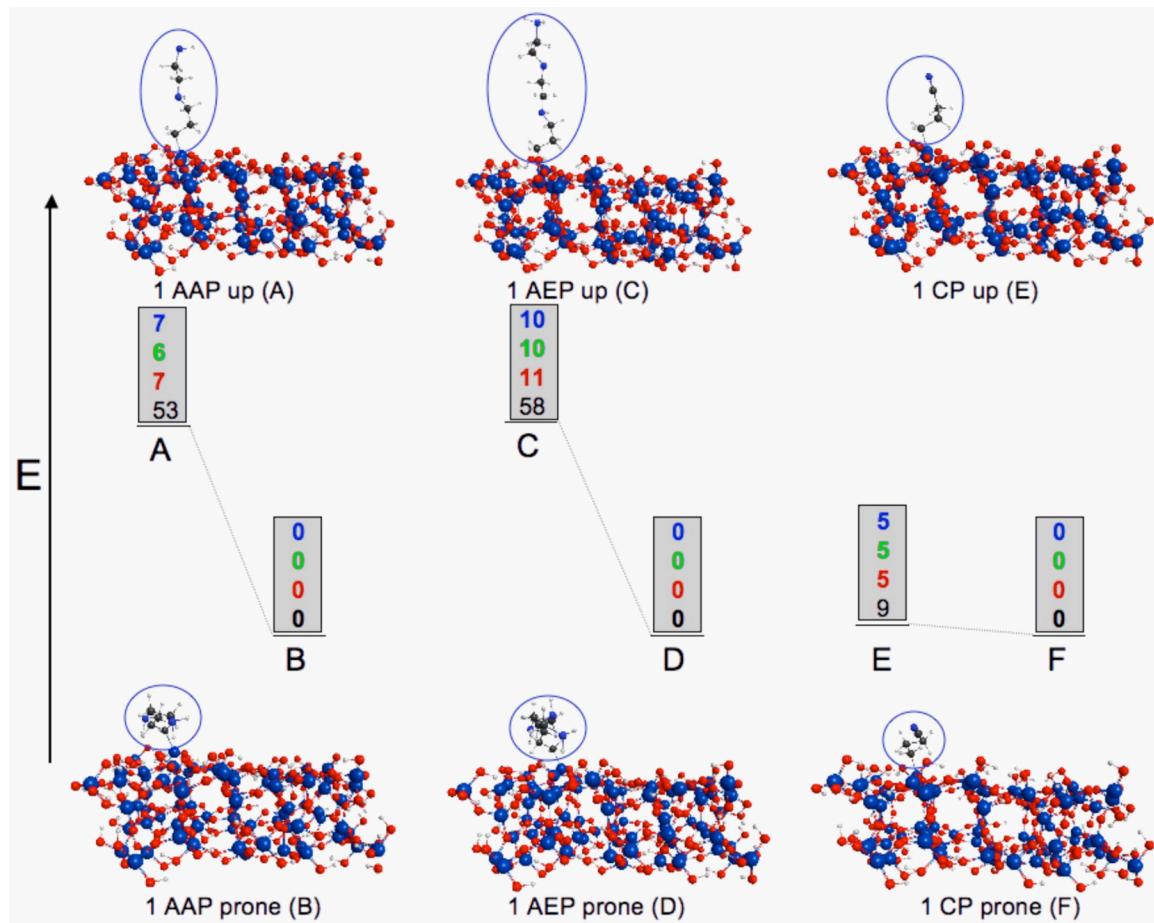
Table 1. Comparison of QM/MM Optimization on Si_6H_{12} /Si_9H_{12} Cluster[a]

	Weiner	SIMOMM/ Tinker	SIMOMM-Rx	GVB-PP(1)	GVB-PP(1)	MM (Tinker)	MM (ReaxFF)
RSM (QM)	Si_6H_{12}	Si_6H_{12}	Si_6H_{12}	Si_6H_{12}	Si_9H_{12}	None	None
BM (MM)	Si_9H_{12}	Si_9H_{12}	Si_9H_{12}	None	None	Si_9H_{12}	Si_9H_{12}
dist (Å)							
2-1	2.278	2.261	2.237	2.159	2.249	2.376	2.360
3-1	2.333	2.338	2.322	2.332	2.329	2.352	2.335
4-1	2.333	2.338	2.322	2.332	2.329	2.352	2.335
5-2	2.345	2.338	2.322	2.332	2.329	2.352	2.335
6-2	2.345	2.338	2.322	2.332	2.329	2.352	2.335
angle (°)							
3-1-2	109.9	108.5	111.9	121.7	106.5	103.9	103.0
4-1-2	109.9	108.5	111.9	121.7	106.5	103.9	103.0
5-2-1	109.9	108.5	111.9	121.7	106.5	103.9	103.0
6-2-1	109.8	108.5	111.9	121.7	106.5	103.9	103.0
3-1-4	112.0	109.3	117.9	116.7	111.0	108.5	97.3
5-2-6	112.0	109.3	117.9	116.7	111.0	108.5	97.3
torsion (°)							
3-1-2-5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4-1-2-3	123.6	118.7	134.9	0.0	118.3	113.4	100.8
6-2-1-5	-123.7	-118.7	-134.9	0.0	-118.3	-113.4	-100.8

[a]The RSM (QM) row indicates the atoms used for the RSM, while the BM (MM) row indicates the whole cluster. The BM contains those atoms that are not in the RSM region. Column one atom numbers used for bond lengths, angles and torsions originate from Figure 2a. Data for columns 2, 3, 5, 6, and 7 were obtained from Ref. 14

Table 2. Comparison of QM/MM Embedding Scheme Optimizations on $Si_6H_{12}/Si_{38}H_{36}$ Hybrid Cluster[a]

	Weiner	SIMOMM/ Tinker	SIMOMM-Rx	GVB-PP(1)	GVB-PP(1)	MM (Tinker)	MM (ReaxFF)
RSM (QM)	Si_6H_{12}	Si_6H_{12}	Si_6H_{12}	Si_6H_{12}	$Si_{38}H_{36}$	None	None
BM (MM)	$Si_{38}H_{36}$	$Si_{38}H_{36}$	$Si_{38}H_{36}$	None	None	$Si_{38}H_{36}$	$Si_{38}H_{36}$
dist (Å)							
2-1	2.277	2.254	2.227	2.159	2.281	2.370	2.363
3-1	2.332	2.342	2.328	2.332	2.349	2.350	2.345
4-1	2.333	2.342	2.328	2.332	2.349	2.350	2.345
5-2	2.346	2.342	2.328	2.332	2.349	2.350	2.345
6-2	2.346	2.342	2.328	2.332	2.349	2.350	2.345
angle (°)							
3-1-2	109.9	108.0	112.6	121.7	105.6	104.3	103.7
4-1-2	109.9	108.0	112.6	121.7	105.6	104.3	103.7
5-2-1	109.9	108.0	112.6	121.7	105.6	104.3	103.7
6-2-1	109.8	108.0	112.6	121.7	105.6	104.3	103.7
3-1-4	112.0	115.7	119.1	116.7	115.9	112.9	106.5
5-2-6	112.0	115.7	119.1	116.7	115.9	112.9	106.5
torsion (°)							
3-1-2-5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4-1-2-3	123.6	125.9	137.9	0.0	123.2	118.6	111.2
6-2-1-5	-123.7	-125.8	-137.9	0.0	-123.2	-118.6	-111.2

[a]The RSM (QM) row indicates the atoms used for the RSM, while the BM (MM) row indicates the whole cluster. The BM contains those atoms that are not in the RSM region. Column one atom numbers used for bond lengths, angles and torsions originate from Figure 2b. Data for columns 2, 3, 5, 6, and 7 were obtained from Ref. 14.

Chapter 4

INCORPORATING A COMPLETELY RENORMALIZED COUPLED CLUSTER APPROACH INTO A COMPOSITE METHOD FOR THERMODYNAMIC PROPERTIES AND REACTION PATHS

A paper published in the Journal of Chemical Physics

Sean A. Nedd, Nathan J. De Yonker, Angela K. Wilson, Piotr Piecuch and Mark S. Gordon

4.1 Abstract.

The correlation consistent composite approach (ccCA), using the S4 complete basis set two-point extrapolation scheme (ccCA-S4), has been modified to incorporate the left-eigenstate completely renormalized coupled cluster method, including singles, doubles and non-iterative triples (CR-CC(2,3)) as the highest level component. The new ccCA-CC(2,3) method predicts thermodynamic properties with an accuracy that is similar to that of the original ccCA-S4 method. At the same time, the inclusion of the single-reference CR-CC(2,3) approach provides a ccCA scheme that can correctly treat reaction pathways that contain certain classes of multi-reference species such as diradicals, which would normally need to be treated by more computationally demanding multi-reference methods. The new ccCA-CC(2,3) method produces a mean absolute deviation (MAD) of 1.7 kcal/mol for predicted heats of formation at 298 K, based on calibration with the G2/97 set of 148 molecules, which is comparable to that of 1.0 kcal/mol obtained using the ccCA-S4 method, while significantly improving the performance of the ccCA-S4 approach in calculations involving more demanding radical and diradical species. Both the ccCA-CC(2,3) and ccCA-S4 composite methods are used to characterize the conrotatory and disrotatory isomerization

pathways of bicyclo[1.1.0]butane to trans-1,3-butadiene, for which conventional coupled cluster methods, such as the CCSD(T) approach used in the ccCA-S4 model and, in consequence, the ccCA-S4 method itself might fail by incorrectly placing the disrotatory pathway below the controtatory one. The ccCA-CC(2,3) scheme provides correct pathway ordering, while providing an accurate description of the activation and reaction energies characterizing the lowest-energy conrotatory pathway. The ccCA-CC(2,3) method is thus a viable method for the analyses of reaction mechanisms that have significant multi-reference character, and presents a generally less computationally intensive alternative to true multi-reference methods, with computer costs and ease of use that are similar to those that characterize the more established, CCSD(T)-based, ccCA-S4 methodology.

4.2 Introduction

Composite electronic structure methods employ several levels of theory in a single series of calculations to predict equilibrium properties of molecules, and they are frequently used to provide thermodynamic properties within chemical accuracy (±1.0 kcal/mol) at a modest computational cost. Composite methods can achieve a relatively low computational cost by taking advantage of the near additivity of improvements in the level of theory and of the atomic basis set.[1-3] Among the composite methods that take advantage of this additivity are the Gaussian (Gn) methods (G1, G2, G3, G4, G3(MP2), and G4(MP2))[4-7], the Weizmann (Wn) methods (W1, W2, W3, W4)[8-10], the Bond Additivity Correction (BAC) methods (BAC-G2, G3B3, G3MP2B3)[11,12], the complete basis set method developed by Petersson and co-workers[13-17], the *ab initio* thermochemical method by Stanton and co-workers[18-20], the Feller-Peterson-Dixon (FPD) composite method[21-26], the multicoefficient correlation method

by Truhlar and co-workers[27-30], and the correlation consistent Composite Approach (ccCA) methods (ccCA-S, ccCA-P, ccCA-S4).[31-34] De Yonker et al[31] have shown that the ccCA method delivers improvements over the more commonly used Gn and Wn series of methods since the ccCA method does not use empirical parameters, in contrast to the Gn methods, and is less computationally demanding than the Wn methods. The Gn methods introduce higher level corrections (HLCs), fitted parameters that are functions of the number of alpha and beta electrons in the system of interest. The total energy corrections from the HLCs can cause a deterioration in the accuracy of the calculated final energies since the magnitude of the HLC corrections increases with system size.[4] The Wn methods include coupled cluster (CC) levels as high as five-particle cluster operators (CCSDTQ5), making the methods very computationally intensive and therefore difficult to apply to molecules with more than a small number of atoms. The ccCA method makes use of correlation consistent basis sets[35-38] that converge systematically to the complete basis set (CBS) limit by using an extrapolation scheme.[31] The ccCA method is more computationally efficient than the Wn methods, since the highest level of theory employed for the ccCA-S4[31] method is coupled cluster with singles, doubles and perturbative triples (CCSD(T)[39]).

In the ccCA-S4 method, energy corrections account for Douglas-Kroll scalar relativistic effects, core-valence corrections, and complete basis set (CBS) two-point extrapolation.[31] The ccCA-S4 method will be discussed in more detail below.

The composite methods mentioned above are all based on single-reference (SR) levels of theory. There are some composite methods that explicitly include multi-reference (MR) levels of theory, including the MR versions of the Gn methods[40], the MR W1CAS and W2CAS methods[41], and the MR-ccCA methods.[42-44] A disadvantage of using explicitly MR

methods is that active spaces can be difficult to choose, and MR methods based on the complete-active-space reference spaces scale exponentially with the size of the system.

The left eigenstate completely renormalized coupled cluster singles, doubles and non-iterative triples (CR-CC(2,3)) approach has been shown to accurately treat single bond breaking and bond formation, as well as chemical reaction profiles involving closed-shell, radical, and diradical species at a computational cost similar to that of CCSD(T).[45,46] Therefore, the SR CR-CC(2,3) method can account for the MR nature of many chemically important systems and can also potentially be incorporated into a composite method to provide accurate treatments of chemical reaction profiles and at least some diradical species. In this paper, both experimental results and the original ccCA-S4 method are employed as calibrations to assess the CR-CC(2,3)-based modification of the method. The new ccCA-CC(2,3) method has been implemented in the GAMESS (General Atomic and Molecular Electronic Structure System)[48] computational package, using the spin-free, RHF-based CR-CC(2,3) routines for closed shells described in Ref. 45 and the spin-integrated, ROHF-based CR-CC(2,3) routines for open shells described in Ref. 46. A spin restricted RHF/ROHF-based ccCA-S4 method, in which the closed-shell calculations are performed with the help of the spin-free CCSD(T) routines described in Ref. 47 and in which the open-shell calculations rely on the ROHF-based CR-CC(2,3) codes,[46] has also been implemented in GAMESS. The aim of the ccCA-CC(2,3) method is to provide thermodynamic properties for closed-shell species and diradicals, as well as for open-shell systems, with an accuracy that is close to the chemical accuracy (± 1.0 kcal/mol mean absolute deviation) provided by the ccCA-S4 method if systems under consideration do not include diradical and other MR species. The ccCA-S4 and ccCA-CC(2,3) methods are compared using the G2/97 set of 148 molecules[5,49],

and by examining the thermal pericyclic rearrangement of bicyclo[1.1.0]butane (bicbut) to trans-buta-1,3-diene, which contains a disrotatory transition state that is a diradical species. The isomerization of bicyclo[1.1.0]butane (bicbut) to trans-buta-1,3-diene presents a good test for the viability of the ccCA-CC(2,3) method, since it is generally difficult to obtain a highly accurate description of the experimentally accessible conrotatory pathway that characterizes this rearrangement reaction, while the higher-energy disrotatory pathway is poorly described by conventional SR approaches, such as CCSD(T).[50] A viable composite method should predict that the conrotatory transition state is favored, and the ability of the ccCA-CC(2,3) approach to do this is examined in the present study.

In the present work, the following ccCA calculations are presented:

(1) The GAMESS RHF/ROHF-based ccCA-S4 (GROS4) method, in which closed-shell calculations use CCSD(T) and open-shell calculations use CR-CC(2,3), is compared with unpublished RHF/ROHF-based ccCA-S4 (uROS4) data, employing CCSD(T) for closed- and open-shells, and experimental heats of formation for the G2/97 set.

(2) The ccCA-CC(2,3) method, also abbreviated as GROS4-CC(2,3), is calibrated against the uROS4 and experimental heats of formation data.

(3) The ccCA-S4 and ccCA-CC(2,3) methods are compared for the structures and relative energies along the thermal pericyclic rearrangement reaction pathways.

The Heats of Formation results from the calibrations among GROS4, GROS4-CC(2,3), and uROS4 are expected to be similar for each species. However, mean absolute deviation (MAD) values will determine the ultimate accuracy for each ccCA calculation of GROS4, GROS4-CC(2,3), uROS4, and also experimental values.

4.3 Theory and computational details.

4.3.1 Description of ccCA theories.

ccCA-S4 method. The ccCA-S4[31,32] method makes use of second order Møller- Plesset perturbation theory (MP2[51]), density functional theory (DFT) using the B3LYP functional[52], and coupled cluster theory with singles, doubles and perturbative triples (CCSD(T))[39]. The notation 'E(A/B)' indicates an energy calculation using method 'A' with the basis set 'B'. The basis sets used in the method are primarily the correlation consistent cc-pVXZ[36,38] and aug-cc-pVXZ[53,54] basis sets. The 6-31G(2df,p) Pople basis set[55] is used for geometry optimizations.[31,32] Scalar relativistic effects are accounted for with the second order Douglas-Kroll-Hess (DK)-derived correlation consistent basis sets, cc-pVXZ-DK[56-58]; core valence (CV) correlation effects are included using the aug-cc-pCVXZ[53,54,59] basis sets.

Energy and structure optimizations, and semi-numerical (finite differencing of analytic first derivatives) Hessian calculations are prepared using B3LYP/6-31G(2df,p) with Cartesian Gaussian functions (6d, 10f).[60,61] All other calculations are single point energy calculations which make use of spherical Gaussian functions (5d, 7f). The following shorthand notation is used:

AVXZ = aug-cc-pVXZ

VTZ-DK = cc-pVTZ-DK (no tight d functions)

NVTZ = cc-pVTZ (no tight d functions)

VTZ = cc-pVTZ (tight d functions on Al − Ar atoms)

ACVTZ = aug-cc-pCVTZ.

The final ccCA-S4 energy at 0K (E(ccCA-S4)) is

$$E(ccCA - S4) = E(MP2/CBS) + \Delta E(SR - MP2) + \Delta E(CC) + \Delta E(CV) + ZPE \quad (1)$$

In Eq. (1) ΔE(SR-MP2), ΔE(CC), and ΔE(CV) are, respectively, the corrections (assumed to be additive) for Douglas-Kroll, for the coupled cluster contributions beyond MP2, and for the core and core-valence correlation effects. ZPE refers to the scaled (0.9854) zero point vibrational energy correction.[55]

The ccCA-S4 energy at 298K (E(ccCA-S4, 298K)) is

$$E(ccCA - S4, 298K) = E(ccCA - S4) + ETH. \tag{2}$$

In Eq. (2) ETH represents the thermal energy correction at 298K. The scaled ZPE and ETH thermodynamic energy term values are obtained from the GAMESS Hessian calculation.

E(MP2/CBS) is the energy resulting from the two point complete basis set extrapolation of E(MP2/AVQZ) and E(MP2/AVTZ) based on the equation[32],

$$E(l_{max}) = E(MP2/CBS) + \frac{B}{\left(l_{max} + \frac{1}{2}\right)^4}, \tag{3}$$

where l_{max} is the highest angular momentum in the basis set and B is a constant. For E(MP2/AVTZ), l_{max} is 3 and for E(MP2/AVQZ), l_{max} is 4; Total energies are used. Using Eq. (3) and the l_{max} value for E(MP2/AVTZ),

$$E(MP2/AVTZ) = E(3) = E(MP2/CBS) + \frac{2^4 \times B}{7^4} \tag{4}$$

and for E(MP2/AVQZ),

$$E(MP2/AVQZ) = E(4) = E(MP2/CBS) + \frac{2^4 \times B}{9^4}. \tag{5}$$

After rearranging equations (4) and (5) to solve for E(MP2/CBS), one obtains

$$E(MP2/CBS) = \left(\frac{6561}{4160}\right) \times E(MP2/AVQZ) - \left(\frac{2401}{4160}\right) \times E(MP2/AVTZ). \tag{6}$$

The ccCA-S4 molecular energy values from equations Eqs. (1) and (2) are used to obtain heats of formation values. The following standard shorthand notation is used for all heats of formation calculations:

ΔH_{f0} = ccCA Heats of Formation at 0K

ΔH_{f298} = ccCA Heats of Formation at 298K

E_{tot0} = sum of atom ccCA total energies at 0K

E_{tot298} = sum of atom ccCA total energies at 298K

ΔH_{a0} = sum of heats of atomization[62-64] at 0K

ΔH_{a298} = sum of heats of atomization[62-64] at 298K

The ΔH_{f0} and ΔH_{f298} for ccCA-S4 are expressed as

$$\Delta H_{f0} = E(ccCA - S4) - E_{tot0} + \Delta H_{a0}, \tag{7}$$

$$\Delta H_{f298} = E(ccCA - S4) - E_{tot298} + \Delta H_{a298}. \tag{8}$$

ccCA-CC(2,3) method. This composite method follows the ccCA-S4 approach described above, but replaces CCSD(T) by CR-CC(2,3). The final ccCA-CC(2,3) energy at 0K (E(ccCA-CC(2,3))) is

$$E(ccCA - CC(2,3)) = E(MP2/CBS) + \Delta E(SR - MP2) + \Delta E(CC) + \Delta E(CV) + ZPE \tag{9}$$

and the final ccCA-CC(2,3) energy at 298K (E(ccCA-CC(2,3) 298K)) is

$$E(ccCA - CC(2,3)\ 298K) = E(ccCA - CC(2,3)) + ETH, \tag{10}$$

where $\Delta E(CC)$ now represents the CR-CC(2,3) energy correction beyond MP2 and is

$$\Delta E(CC) = E(CR - CC(2,3)/VTZ) - E(MP2/VTZ). \tag{11}$$

The ccCA-CC(2,3) heats of formation are calculated in the same way as for the ccCA-S4:

$$\Delta H_{f0} = E(ccCA - CC(2,3)) - E_{tot0} + \Delta H_{a0} \tag{12}$$

and ΔH_{f298} for ccCA-CC(2,3) is

$$\Delta H_{f298} = E(ccCA - CC(2,3)) - E_{tot298} + \Delta H_{a298}. \tag{13}$$

The calibration comparisons, presented in the following paragraphs, compare the GAMESS RHF/ROHF-based ccCA-S4 (GROS4) and ccCA-CC(2,3) predictions, where GROS4 uses RHF-based CCSD(T) for closed shells and ROHF-based CR-CC(2,3) for open shells and ccCA-CC(2,3) employs the restricted CR-CC(2,3) approach for both closed and open shells, with those from the unpublished RHF/ROHF-based ccCA-S4 (uROS4) calculations, in which CCSD(T) is utilized throughout, and experimental heats of formation. These comparisons employ the 298K heats of formation from the ccCA-S4 and ccCA-CC(2,3) methods. The ccCA-S4 and ccCA-CC(2,3) methods are also used to obtain the relative energies of structures along the thermal pericyclic rearrangement reaction pathway described below.

4.3.2 Thermal pericyclic rearrangement calculations. The thermal pericyclic rearrangement follows two pathways: a disrotatory pathway and a conrotatory pathway. Each of the pathways, as shown in Figure 1, contains five species, four of which are the same for both pathways. The one species that differs is the initial transition state structure. The initial transition state is either of the conrotatory type (for a conrotatory pathway) or of the disrotatory type (for a disrotatory pathway). These pericyclic rearrangement pathways are parts of concerted mechanisms in which the conrotatory pathway appears to be favored.[50,65-70] All six species in the two reaction paths are illustrated in Figure 1, with the following shorthand notation in parentheses: bicyclo[1.1.0]butane (bicbut) reactant, two transition states (disrotatory transition state (dis_TS) and conrotatory transition state (con_TS)),

gauche-buta-1,3-diene (g-but) local minimum, gauche transition state (gt_TS), and *trans*-buta-1,3-diene final product (t-but).

The energies of all six structures were determined using the ccCA-S4 and ccCA-CC(2,3) methods implemented in GAMESS, as described above. Structures for each of the species in the pericyclic rearrangement, as well as the corresponding zero-point vibrational energies, were taken from supporting material provided in a previous paper.[50] These structures and zero-point energies were obtained using the complete active space self-consistent field (CASSCF) approach,[71,72] employing 10 active orbitals and 10 active electrons, and referred to as CASSCF(10,10). These CASSCF(10,10) based geometries replace the prescribed B3LYP/6-31G(2df,p) based geometries in the ccCA-S4 and ccCA-CC(2,3) methods, since it is established in Ref. 50 that B3LYP-based geometries are insufficient for describing the pericyclic rearrangement.

4.4 Results and Discussion.

4.4.1 ccCA-S4 and ccCA-CC(2,3) calibration and comparisons. Tables 1 - 4 list 298K heats of formation for the ccCA-S4 and ccCA-CC(2,3) methods and compare their values with those obtained from experiment (expt.), as well as unpublished RHF/ROHF-based ccCA-S4 (uROS4) results. The mean absolute deviations (MAD) are included in these tables.

In Table 1, for the G2-1 set (the first 55 species of the G2/97 set) the MAD between ccCA-S4 in GAMESS (GROS4) and experiment is 0.8 kcal/mol; that between GROS4 and uROS4 is 0.5 kcal/mol; and that between uROS4 and experiment is 0.8 kcal/mol. So, for the G2-1 set, the ccCA-S4 method is consistent with uROS4 data. In general, a measure of good chemical accuracy is a MAD value less than 1.0 kcal/mol. Since the CR-CC(2,3) coupled

cluster method is used as a substitute for CCSD(T) in the GROS4 implementation of ccCA-S4 for open-shell species, the agreement between GROS4 and its uROS4 counterpart which relies on CCSD(T) for both closed and open shells shown in Table 1 is encouraging. In order to directly compare CCSD(T) based results, the singlet species MAD values for the G2-1 set are also presented in Table 1. For the ccCA-S4 method, the G2-1 singlet species MAD between GROS4 and experiment is 0.6 kcal/mol; that between GROS4 and uROS4 is 0.4 kcal/mol; and that between uROS4 and experiment is 0.8 kcal/mol. These MAD values for the G2-1 singlet species indicate that the GROS4 method does provide heats of formation that are within chemical accuracy, overall.

Table 2 contains a similar comparison to Table 1 for the G2-2 set, which represents the remaining 97 species in the full G2 set. For ccCA-S4 the MAD value for the G2-2 set between GROS4 and experiment is 1.1 kcal/mol; that between GROS4 and uROS4 is 0.8 kcal/mol; and that between uROS4 and experiment is 0.8 kcal/mol. The higher MAD values are a reflection of larger errors in the both the singlets and non-singlets in the G2-2 set. For example, the MAD values between GROS4 and experiment for just the G2-2 singlet species is 1.1 kcal/mol, an increase of 0.5 kcal/mol relative to the G2-1 set, while the MAD value for the G2-2 non singlet species is 1.5 kcal/mol. Overall, for the entire G2/97 set (G2-1 + G2-2) the MAD between GROS4 and experiment is 1.0 kcal/mol; that between GROS4 and uROS4 is 0.7 kcal/mol; and that between uROS4 and experiment is 0.8 kcal/mol, all within chemical accuracy. For ccCA-S4, the small changes in MAD errors between GROS4 and uROS4 for all G2/97 species (0.2 kcal/mol) and for only the G2/97 singlet species (0.1 kcal/mol) indicate that using the ROHF-based CR-CC(2,3) approach is an acceptable substitution for using the ROCCSD(T) method. The maximum (MAX) error between GROS4 and

experiment for all species in the G2/97 set is 4.6 kcal/mol, which is only 0.4 kcal/mol greater than the corresponding MAX error between uROS4 and experiment. The MAX error between GROS4 and uROS4 for all species in the G2/97 set is lower at a value of 3.2. The small MAX errors also indicate, together with the small MAD errors, that the GAMESS ccCA-S4 method provides acceptable results.

In Table 3, for all species in the G2-1 set using the ccCA-CC(2,3) method, in which CR-CC(2,3) is consistently used for closed- as well as for open-shell species (designated in Table 3 as GROS4-CC(2,3)), the MAD between the GROS4-CC(2,3) data and experiment is 0.9 kcal/mol (0.1 kcal/mol greater than the MAD from ccCA-S4); that between GROS4-CC(2,3) and uROS4 is 0.9 kcal/mol; and that between uROS4 and experiment is 0.8 kcal/mol. For only singlet species, the MAD between GROS4-CC(2,3) and experiment is 0.9 kcal/mol (0.3 kcal/mol greater than the ccCA-S4 MAD); that between GROS4-CC(2,3) and uROS4 is 1.0 kcal/mol; and that between uROS4 and experiment was 0.8 kcal/mol, all within chemical accuracy.

The ccCA-CC(2,3) MAD value for the G2-2 set (Table 4), comparing GROS4-CC(2,3) with experiment, is 2.2 kcal/mol; that between GROS4-CC(2,3) and uROS4 is 2.2 kcal/mol; and that between uROS4 and experiment is 0.8 kcal/mol. The higher MAD values are a reflection of larger errors for both the singlets and non-singlets in the G2-2 set, as shown above for ccCA-S4. For example, for ccCA-CC(2,3), the MAD value relative to experiment for the G2-2 singlet species is 2.4 kcal/mol, which is an increase of 1.5 kcal/mol in MAD error from the ccCA-CC(2,3) G2-1 set values, and reflects the more complex structures that are considered in the G2-2 set. On the other hand, the MAD between GROS4-CC(2,3) and experiment, for the G2-2 non singlet species is 0.9 kcal/mol. This is an

improvement over the ccCA-S4 G2-2 non singlet species by 0.6 kcal/mol. Overall, Table 4 indicates that for ccCA-CC(2,3) the G2/97 set MAD between GROS4-CC(2,3) and experiment is 1.7 kcal/mol; that between GROS4-CC(2,3) and uROS4 is 1.7 kcal/mol; and that between uROS4 and experiment is 0.8 kcal/mol. The ccCA-CC(2,3) method improves on the predicted heats of formation of the doublet species, for example, for NO_2 the absolute difference from experiment for the ccCA-S4 heats of formation is 4.0 kcal/mol vs. 1.7 kcal/mol for ccCA-CC(2,3). For $(CH_3)_3C$, the analogous improvement is 2.4 kcal/mol (ccCA-S4) to 0.5 kcal/mol (ccCA-CC(2,3)).

The MAX error for all species in the G2/97 set between GROS4-CC(2,3) and experiment increased by 2.6 kcal/mol compared to the same MAX error between GROS4 and experiment. Also, the MAX error for all species in the G2/97 set between GROS4-CC(2,3) and uROS4 increased by 3.3 kcal/mol compared to that between GROS4 and uROS4. The discussion below discusses these acceptable increases.

GROS4 and uROS4 may appear to be the same for certain closed shell species, but this does not imply that the energies of atomization will be the same. For example, if a closed-shell species AB dissociates into open-shell A and B fragments, the GROS4 energy is calculated using CCSD(T)/RHF for AB and CR-CC(2,3)/ROHF for A and B. The analogous uROS4 energy uses CCSD(T)/RHF for AB and CCSD(T)/UHF for A and B. So, the energies of atomization will be different despite the fact that the total energy of AB may be the same in both calculations. Similarly, the fact that GROS4 and GROS4-CC(2,3) may be the same for open shells does not mean that the atomization energies are the same, since one of the dissociation fragments could be a closed shell (so that GROS4 would use CCSD(T)/RHF and GROS4-CC(2,3) would use CR-CC(2,3)/RHF for this fragment). Only if the molecule and

fragments are all open shells, should GROS4 and GROS4-CC(2,3) be expected to give the same atomization energy, since all species involved in atomization are treated by CR-CC(2,3)/ROHF (see, e.g., NH_2 in Table 3).

4.4.2 Pericyclic rearrangement. Figure 2 illustrates the potential energy curves for the pericyclic rearrangement. The only available experimental data are for the conrotatory transition state and the *trans*-buta-1,3-diene final product. The activation energy barrier for the isomerization bicyclo[1.1.0]butane into buta-1,3-diene is 40.6±2.5 kcal/mol at 0K,[73] and since this isomerization is a conrotatory process, the activation energy corresponds to the conrotatory transition state. The *trans*-buta-1,3-diene final product has a heat of reaction that is estimated to be -25.9±0.4 kcal/mol at 298K.[74] Thus, only the conrotatory transition state and *trans*-buta-1,3-diene final product energies can be compared to experimental data. This lack of experimental data for the pericyclic rearrangement pathway also emphasizes the need for the development of quantitatively reliable methods to predict the energies along reaction pathways that contain diradical species. It is therefore gratifying that all conrotatory activation energies and the net reaction energy predicted by ccCA-S4 and ccCA-CC(2,3) agree to within 1-2 kcal/mol with the experimental values (see Figure 2 and Table 5).

A simple measure of the percent diradical character in a wavefunction is $(2 - n_H)$ x 100, where n_H is the occupation number of the natural orbital corresponding to the highest occupied molecular orbital (HOMO). In this work, n_H was obtained from the CASSCF(10,10)/cc-pVDZ wavefunction.[50] The percent diradical character for each of the relevant species in the pericyclic reaction mechanism is listed in Table 5. One advantage of the CR-CC(2,3) method is that, unlike traditional coupled cluster approaches, such as CCSD(T), CR-CC(2,3) can correctly account for significant diradical character. This

capability of the CR-CC(2,3) approach is relevant in the current context, since the various species listed in Table 5 have dramatically different amounts of diradical character. So, for any method to be successful for this system, it must be able to account correctly for diradical character that varies a great deal along the reaction potential energy surface. For example, the diradical character in the conrotatory vs. the disrotatory transition state is 24% vs. 90%, while the diradical character in the local minimum energy structures is less than 10%. As may be seen in Table 5, this leads to large differences in predicted relative energies between CR-CC(2,3) and CCSD(T), with CCSD(T) placing the disrotatory transition state below the conrotatory one, contradicting the well-established experimental data[65,66] and MR-[50,67,69] or MR-based Quantum Monte Carlo[70] calculations. The agreement between these two levels of theory is much better for the conrotatory than for the disrotatory transition state, due to the smaller diradical character in the former than in the latter. The above accuracy patterns characterizing the restricted CR-CC(2,3) and CCSD(T) methods propagate into the corresponding CR-CC(2,3)-based ccCA-CC(2,3) and CCSD(T)-based ccCA-S4 calculations, where only the former ccCA scheme provides the correct energy ordering of both isomerization pathways.

The ccCA-S4 and ccCA-CC(2,3) values of the reaction heat at 298K were also calculated in order to compare with the available experimental reaction enthalpy, which is, as mentioned above, -25.9±0.4 kcal/mol at 298K.[74] The ccCA-S4 and ccCA-CC(2,3) approaches give -26.7 kcal/mol and -26.6 kcal/mol, respectively, using the CASSCF(10,10)/cc-pVDZ geometries. These ccCA heats of reaction are in very good agreement with experiment and with each other.

4.5 Conclusions

Both the ccCA-S4 and ccCA-CC(2,3) methods have been compared with the G2/97 set using a new code that has been developed for composite methods. The ROHF-based CR-CC(2,3) method has been shown to be an accurate and computationally cost effective option for the calculation of heats of formation, both within the ccCA-CC(2,3) approach, in which the CR-CC(2,3) method is used for closed- as well open-shell species, and within the modified ccCA-S4 protocol, in which CCSD(T) is used for closed-shell systems and CR-CC(2,3) is a substitute for CCSD(T) in calculations involving open-shell species. The ccCA-CC(2,3) method, gives a MAD of 1.7 kcal/mol relative to experiment for the G2/97 set. For open shell species and diradicals, which can frequently be problematic for composite methods, the ccCA-CC(2,3) MAD error is only 0.9 kcal/mol relative to experiment. These results make the new method competitive with the successful and well-established ccCA-S4 approach. Relative to experiment, for open shell species and diradicals, the ccCA-CC(2,3) method improves on the ccCA-S4 method by a MAD error of 0.6 kcal/mol. This ability to reliably treat open shell and diradical species manifests itself in the correct resolution by the ccCA-CC(2,3) method of the competing reaction pathways for the pericyclic rearrangement of bicyclo[1.1.0]butane to trans-1,3-butadiene.

Since the ccCA-CC(2,3) method is able to consistently predict the relative energies for systems that contain species of high diradical character, and for potential energy surfaces in which the diradical character changes significantly from structure to structure, it is possible that other composite methods can take advantage of the CR-CC(2,3) approach as a viable replacement for the CCSD(T) method. In particular, composite methods that employ CR-CC(2,3) can potentially treat systems with a more significant MR character, without the

need for true MR methods that are computationally demanding and sometimes require difficult choices when determining an active space.

4.7 References.

[1] L. A. Curtiss, J. E. Carpenter, K. Raghavachari, and J. A. Pople, J. Chem. Phys. **96**, 9030 (1992).

[2] M. L. McKee and W. N. Lipscomb, J. Am. Chem. Soc. **103**, 4673 (1981).

[3] E. W. Ignacio and H. B. Schlegel, J. Comp. Chem. **12**, 751 (1991).

[4] J. A. Pople, M. Head-Gordon, and D. J. Fox, J. Chem. Phys. **90**, 5622 (1989).

[5] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. **94**, 7221 (1991).

[6] L. A. Curtiss, P. C. Redfern, and K. Raghavachari, J. Chem. Phys. **126**, 084108 (2007).

[7] L. A. Curtiss, P. C. Redfern, and K. Raghavachari, J. Chem. Phys. **127**, 124105 (2007).

[8] S. Parthiban and J. M. L. Martin, J. Chem. Phys. **114**, 6014 (2001).

[9] A. D. Boese, M. Oren, O. Atasoylu, and J. M. L. Martin, J. Chem. Phys. **120**, 4129 (2004).

[10]A. Karton, E. Rabinovich, and J. M. L. Martin, J. Chem. Phys. **125**, 144108 (2006).

[11]C. F. Melius and M. D. Allendorf, J. Phys. Chem. **104**, 2168 (2000).

[12]B. Anantharaman and C. F. Melius, J. Phys. Chem. **94**, 1734 (2005).

[13]G. A. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al-Laham, W. A. Shirley, and J. Mantzaris, J. Chem. Phys. **89**, 2193 (1988).

[14]G. A. Petersson and M. A. Al-Laham, J. Chem. Phys. **94**, 6081 (1991).

[15]G. P. F. Wood, L. Radom, G. A. Petersson, E. C. Barnes, M. J. Frisch, and J. J. A. Montgomery, J. Chem. Phys. **125**, 094106 (2006).

[16]G. A. Petersson, T. G. Tensfeldt, and J. J. A. Montgomery, J. Chem. Phys. **94**, 6091 (1991).

[17]J. J. A. Montgomery, J. W. Ochterski, and G. A. Petersson, J. Chem. Phys. **101**, 5900 (1994).

[18]A. Tajti, P. G. Szalay, A. G. Csaszar, M. Kallay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vazquez, and J. F. Stanton, J. Chem. Phys. **121**, 11599 (2004).

[19]Y. J. Bomble, J. Vazquez, M. Kallay, C. Michauk, P. G. Szalay, A. G. Csaszar, J. Gauss, and J. F. Stanton, J. Chem. Phys. **125**, 064108 (2006).

[20]M. E. Harding, J. Vazquez, B. Ruscic, A. K. Wilson, J. Gauss, and J. F. Stanton, J. Chem. Phys. **125**, 114111 (2008).

[21]D. A. Dixon, D. Feller, and G. Sandrone, J. Phys. Chem. A **103**, 4744 (1999).

[22]D. Feller and D. A. Dixon, J. Phys. Chem. A **103**, 6413 (1999).

[23]D. Feller and D. A. Dixon, J. Phys. Chem. A **104**, 3048 (2000).

[24]D. Feller, D. A. Dixon, and J. S. Francisco, J. Phys. Chem. A **107**, 1604 (2003).

[25]D. Feller, K. A. Peterson, W. A. de Jong, and D. A. Dixon, J. Chem. Phys. **118**, 3510 (2003).

[26]L. Pollack, T. L. Windus, W. A. de Jong, and D. A. Dixon, J. Phys. Chem. A **109**, 6934 (2005).

[27]P. L. Fast, J. Corchado, M. L. Sanchez, and D. G. Truhlar, J. Phys. Chem. A **103**, 3139 (1999).

[28]P. L. Fast, J. C. Corchado, M. L. Sanchez, and D. G. Truhlar, J. Phys. Chem. A **103**, 5129 (1999).

[29]P. L. Fast and D. G. Truhlar, J. Phys. Chem. A **104**, 6111 (2000).

[30]P. L. Fast, N. E. Schultz, and D. G. Truhlar, J. Phys. Chem. A **105**, 4143 (2001).

[31]N. J. DeYonker, T. Grimes, S. Yockel, A. A. Dinescu, B. Mintz, A. K. Wilson, and T. R. Cundari, J. Chem. Phys. **125**, 104111 (2006).

[32]N. J. De Yonker, T. R. Cundari, and A. K. Wilson, J. Chem. Phys. **124**, 114104 (2006).

[33]N. J. De Yonker, B. Mintz, T. R. Cundari, and A. K. Wilson, J. Chem. Theor. Comput. **4**, 328 (2008).

[34]N. J. De Yonker, B. R. Wilson, A. W. Pierpont, T. R. Cundari, and A. K. Wilson, Molecular Physics **107**, 1107 (2009).

[35]T. H. Dunning Jr., J. Chem. Phys. **90**, 1007 (1989).

[36]R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

[37]D. E. Woon and T. H. Dunning Jr., J. Chem. Phys. **98** , 1358 (1993).

[38]D. E. Woon and T. H. Dunning Jr., J. Chem. Phys. **100**, 2975 (1994).

[39]K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[40]T. I. Sølling, D. M. Smith, L. Radom, M. A. Freitag, and M. S. Gordon, J. Chem. Phys. **115**, 8758 (2001).

[41] J. M. L. Martin and S. Parthinban, in *Quantum Mechanical Prediction of Thermochemical Data*, edited by J. C. a. A. Szarecka (Kluwar Academic Publishers, Dordrecht, The Netherlands, 2001), Vol. 22, pp. 31.

[42] B. Mintz, T. G. Williams, L. Howard, and A. K. Wilson, J. Chem. Phys. **130**, 234104 (2009).

[43] G. A. Oyedepo, C. Peterson, and A. K. Wilson, J. Chem. Phys. **135**, 094103 (2011).

[44] W. Jiang and A. K. Wilson, J. Chem. Phys. **134**, 034101 (2011).

[45] P. Piecuch and M. Wloch, J. Chem. Phys. **123**, 224105 (2005).

[46] M. Wloch, J. R. Gour, and P. Piecuch, J. Phys. Chem. A **111**, 11359 (2007).

[47] P. Piecuch, S. A. Kucharski, K. Kowalski, and M. Musiał, Comp. Phys. Comm. **149**, 71 (2002).

[48] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, and N. Matsunaga, Journal of Computational Chemistry **14**, 1347–1363 (1993).

[49] L. A. Curtiss, P. C. Redfern, K. Raghavachari, V. Rassalov, and J. A. Pople, J. Chem. Phys. **110**, 4703 (1999).

[50] A. Kinal and P. Piecuch, J. Phys. Chem. A **111**, 734 (2007).

[51] C. Moller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).

[52] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[53] D. E. Woon and J. Dunning, T. H., J. Chem. Phys. **103**, 4572 (1995).

[54] K. A. Peterson and T. H. Dunning Jr., J. Chem. Phys. **117**, 10548 (2002).

[55] L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, J. Chem. Phys. **114**, 108 (2001).

[56] W. A. de Jong, R. J. Harrison, and D. A. Dixon, J. Chem. Phys. **114**, 48 (2001).

[57]M. Douglas and N. M. Kroll, Ann. Phys. **82**, 89 (1974).

[58]B. A. Hess, Phys. Rev. A **33**, 3742 (1986).

[59]S. Yockel and A. K. Wilson, Theochem **120**, 119 (2008).

[60]S. F. Boys, Proc. R. Soc. (London) A **200**, 542 (1950).

[61]F. Jensen, *Introduction to Computational Chemistry*, 2 ed. (Wiley, Odense, 2007).

[62]L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **106**, 1063 (1997).

[63]A. Karton and J. M. L. Martin, J. Phys. Chem. A **111**, 5936 (2007).

[64] NIST Standard Reference Database 101, 2011.

[65]K. B. Wiberg and J. M. Lavanish, J. Am. Chem. Soc. **88**, 5272 (1966).

[66]G. L. Closs and P. E. Pfeffer, J. Am. Chem. Soc. **90**, 2452 (1968).

[67]K. A. Nguyen and M. S. Gordon, J. Am. Chem. Soc. **117**, 3835 (1995).

[68]J. J. Lutz and P. Piecuch, J. Chem. Phys. **128**, 154116 (2008).

[69]D. A. Mazziotti, J. Phys. Chem. A **112**, 13684 (2008).

[70]R. Berner and A. Lüchow, J. Phys. Chem. A **114**, 13222 (2010).

[71]M. Schmidt and M. S. Gordon, Ann. Rev. Phys. Chem. **49**, 233 (1998).

[72]B. O. Roos, in *Advances in Chemical Physics*, edited by K. P. Lawley (Wiley interscience, New York, 1987), Vol. 69, pp. 339.

[73]R. Srinivasan, A. Levi, and I. Haller, J. Phys. Chem. **58**, 1775 (1965).

[74]K. B. Wiberg and R. A. Fenoglio, J. Am. Chem. Soc. **90**, 3395 (1968).

Figure 1. Pathways in the thermal pericyclic rearrangement (see text for description).
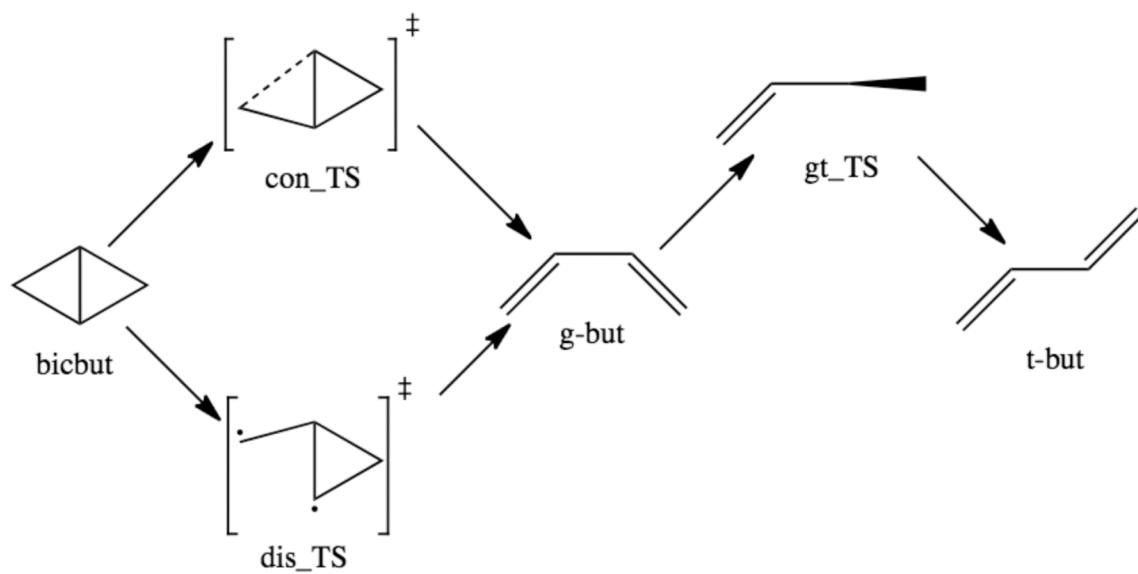
Figure 2. The conrotatory pathway (dashed curve) and disrotatory pathway (solid curve)
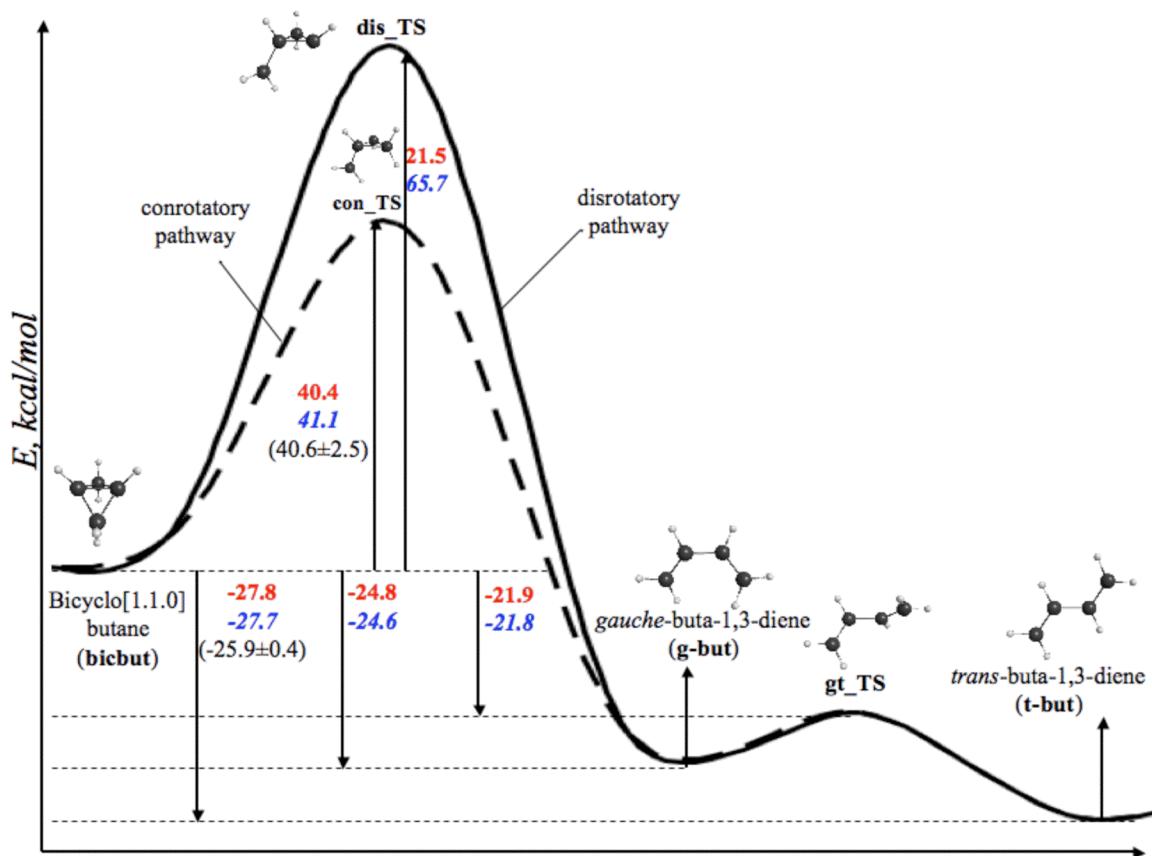
Table 1. ccCA-S4 Heats of Formation and MAD[a] values (kcal/mol) for the G2-1 set compared with experiment (expt.), GROS4, and uROS4.

G2-1 Set		ccCA-S4 ΔH_f (298K), kcal/mol					
Species	Multiplicity	expt. (1)	GROS4 (2)	uROS4 (3)	(2)-(1)	(2)-(3)	(3)-(1)
LiH	1	33.3	32.4	32.4	-0.9	0.0	-0.9
BeH	2	81.7	81.6	81.0	-0.1	0.6	-0.7
CH	2	142.5	142.5	142.3	-0.1	0.1	-0.2
CH$_2$	3	93.7	94.2	94.2	0.5	0.0	0.5
CH$_2$	1	102.8	102.4	102.3	-0.5	0.1	-0.5
CH$_3$	2	35.0	35.0	35.0	0.0	0.0	0.0
CH$_4$	1	-17.9	-17.9	-18.0	0.0	0.1	-0.1
NH	3	85.2	85.7	85.9	0.5	-0.1	0.7
NH$_2$	2	45.1	44.2	44.2	-0.9	0.0	-0.9
NH$_3$	1	-11.0	-11.7	-11.6	-0.7	-0.1	-0.6
OH	2	9.4	8.9	8.6	-0.5	0.2	-0.8
OH$_2$	1	-57.8	-58.8	-59.0	-1.0	0.2	-1.2
FH	1	-65.1	-66.0	-66.1	-0.9	0.0	-1.0
SiH$_2$	1	65.2	64.1	63.0	-1.1	1.1	-2.2
SiH$_2$	3	86.2	87.3	86.1	1.1	1.2	-0.1
SiH$_3$	2	47.9	48.0	46.8	0.1	1.2	-1.1
SiH$_4$	1	8.2	7.5	6.3	-0.7	1.2	-1.9
PH$_2$	2	33.1	32.0	32.4	-1.1	-0.4	-0.7
PH$_3$	1	1.3	1.3	1.5	0.0	-0.2	0.2
SH$_2$	1	-4.9	-5.5	-5.9	-0.6	0.3	-1.0
ClH	1	-22.1	-22.5	-22.6	-0.4	0.1	-0.5
Li$_2$	1	51.6	50.5	50.5	-1.1	0.0	-1.1
LiF	1	-80.1	-81.9	-82.0	-1.8	0.1	-1.9
C$_2$H$_2$	1	54.2	55.4	55.2	1.2	0.3	1.0
C$_2$H$_4$	1	12.5	12.7	12.5	0.2	0.2	0.0
C$_2$H$_6$	1	-20.1	-20.3	-20.4	-0.2	0.1	-0.3

[a] The MAD values are reported between each of the three data sets as explained in the text. MAD G2-1 is the MAD for the whole G2-1 set and MAD singlet species is for all of the singlet species in the G2-1 set.

Table 1. (continued)

CN	2	104.9	107.5	108.2	2.6	-0.7	3.3
HCN	1	31.5	31.6	31.5	0.1	0.2	0.0
CO	1	-26.4	-26.3	-26.8	0.1	0.5	-0.4
HCO	2	10.0	11.3	10.0	1.3	1.2	0.0
H_2CO	1	-26.0	-26.2	-26.5	-0.2	0.4	-0.5
H_3COH	1	-48.0	-48.7	-48.9	-0.7	0.2	-0.9
N_2	1	0.0	0.0	-0.2	0.0	0.2	-0.2
H_2NNH_2	1	22.8	22.1	22.1	-0.7	0.0	-0.7
NO	2	21.6	23.2	22.3	1.6	0.9	0.7
O_2	3	0.0	1.7	1.1	1.7	0.6	1.1
HOOH	1	-32.5	-32.8	-33.1	-0.3	0.3	-0.6
F_2	1	0.0	0.9	0.9	0.9	0.0	0.9
CO_2	1	-94.1	-93.4	-94.0	0.7	0.6	0.1
Na_2	1	34.0	32.5	32.9	-1.5	-0.3	-1.1
Si_2	3	139.9	142.9	140.3	3.0	2.7	0.4
P_2	1	34.3	35.4	35.5	1.1	-0.2	1.2
S_2	3	30.7	32.4	30.9	1.7	1.4	0.2
Cl_2	1	0.0	0.3	0.1	0.3	0.2	0.1
NaCl	1	-43.6	-43.3	-43.2	0.3	-0.1	0.4
SiO	1	-24.6	-22.6	-23.8	2.0	1.2	0.8
SC	1	66.9	67.5	66.8	0.6	0.7	-0.1
SO	3	1.2	3.3	2.1	2.1	1.2	0.9
ClO	2	24.2	27.1	25.8	2.9	1.3	1.6
FCl	1	-13.2	-13.4	-13.6	-0.2	0.2	-0.4
Si_2H_6	1	19.1	18.5	16.1	-0.6	2.3	-3.0
CH_3Cl	1	-19.6	-19.9	-20.1	-0.3	0.2	-0.5
H_3CSH	1	-5.5	-5.9	-6.3	-0.4	0.4	-0.8
HOCl	1	-17.8	-18.6	-18.9	-0.8	0.3	-1.1
SO_2	1	-71.0	-70.2	-70.7	0.8	0.5	0.3
				MAD of G2-1	0.8	0.5	0.8
				MAD singlet species	0.6	0.4	0.8

Table 2. ccCA-S4 Heats of Formation and corresponding MAD[a] values (kcal/mol) for the G2-2 set comparing experiment (expt.), GROS4, and uROS4.

G2-2 Set		ccCA-S4 ΔH_f (298K), kcal/mol					
Species	Multiplicity	expt. (1)	GROS4(2)	uROS4 (3)	(2)- (1)	(2)- (3)	(3)- (1)
BF_3	1	-271.4	-270.9	-269.7	0.5	-1.2	1.7
BCl_3	1	-96.3	-95.9	-95.4	0.4	-0.6	0.9
AlF_3	1	-289.0	-287.1	-289.9	1.9	2.9	-0.9
$AlCl_3$	1	-139.7	-138.0	-140.9	1.7	2.9	-1.2
CF_4	1	-223.0	-222.9	-223.1	0.1	0.2	-0.1
CCl_4	1	-22.9	-22.2	-22.9	0.7	0.7	0.0
COS	1	-33.1	-33.4	-34.3	-0.3	0.9	-1.2
CS_2	1	28.0	28.1	27.0	0.1	1.1	-1.0
CF_2O	1	-149.1	-144.5	-144.9	4.6	0.4	4.2
SiF_4	1	-386.0	-383.9	-385.2	2.1	1.2	0.8
$SiCl_4$	1	-158.4	-156.0	-157.5	2.5	1.6	0.9
N_2O	1	19.6	20.2	19.9	0.6	0.3	0.3
ClNO	1	12.4	13.8	13.3	1.4	0.5	0.9
NF_3	1	-31.6	-31.1	-31.2	0.5	0.1	0.4
PF_3	1	-229.1	-227.8	-227.6	1.3	-0.2	1.5
O3	1	34.1	36.6	36.1	2.5	0.6	2.0
F_2O	1	5.9	7.1	6.9	1.2	0.1	1.0
ClF_3	1	-38.0	-38.4	-38.7	-0.4	0.3	-0.7
C_2F_4	1	-157.4	-160.2	-160.7	-2.8	0.5	-3.3
C_2Cl_4	1	-3.0	-4.4	-5.3	-1.4	0.9	-2.3
CF_3CN	1	-118.4	-117.6	-118.1	0.8	0.5	0.3
Propyne	1	44.2	45.3	44.9	1.1	0.4	0.7
Allene	1	45.5	46.3	46.0	0.8	0.3	0.5
Cyclopropene	1	66.2	68.7	68.4	2.5	0.3	2.2

[a] The MAD values are reported between each of the three data sets as explained in the text. The five consecutive MAD labels at the end of the table are respectively, the MAD for all species in the G2-2 set, the MAD for all singlet species in the G2-2 set, the MAD for all species in the G2/97 set, the MAD for all singlet species in the G2/97 set, and the MAD for non-singlet species in the G2-2 set. The maximum (MAX) error is given for all species in the G2/97 set.

Table 2. (continued)

Propylene	1	4.8	5.1	4.8	0.3	0.3	0.0
Cyclopropane	1	12.7	13.3	12.9	0.6	0.4	0.2
Propane	1	-25.0	-25.3	-25.5	-0.3	0.2	-0.5
Butadiene	1	26.3	27.8	27.1	1.5	0.6	0.8
2-Butyne	1	34.8	36.6	36.1	1.8	0.5	1.3
Methylene Cyclopropane	1	47.9	47.3	46.8	-0.6	0.5	-1.1
Bicyclobutane	1	51.9	54.7	54.1	2.8	0.6	2.2
Cyclobutene	1	37.4	39.5	39.0	2.1	0.5	1.6
Cyclobutane	1	6.8	7.0	6.5	0.2	0.5	-0.3
Isobutene	1	-4.0	-3.1	-3.9	0.9	0.8	0.1
Transbutane	1	-30.0	-30.0	-30.6	0.0	0.5	-0.6
Isobutene	1	-32.1	-32.0	-32.4	0.1	0.3	-0.3
Spiropentane	1	44.3	45.3	44.7	1.0	0.6	0.4
Benzene	1	19.7	21.8	20.8	2.1	1.0	1.1
CH_2F_2	1	-107.7	-108.0	-108.1	-0.3	0.1	-0.4
CHF_3	1	-166.6	-166.5	-166.7	0.1	0.2	-0.1
CH_2Cl_2	1	-22.8	-22.6	-22.9	0.2	0.3	-0.1
$CHCl_3$	1	-24.7	-23.9	-24.4	0.8	0.5	0.3
Methylamine	1	-5.5	-5.7	-5.8	-0.2	0.1	-0.3
Methyl Cyanide	1	17.7	18.4	18.1	0.7	0.3	0.4
Nitromethane	1	-17.8	-17.2	-17.8	0.6	0.6	0.0
Methyl Nitrite	1	-15.9	-15.5	-16.0	0.4	0.5	-0.1
Methyl Silane	1	-7.0	-6.3	-7.5	0.7	1.2	-0.5
Formic Acid	1	-90.5	-90.4	-91.0	0.1	0.6	-0.5
Methyl Formate	1	-85.0	-84.9	-86.1	0.1	1.2	-1.1
Acetamide	1	-57.0	-55.4	-56.6	1.6	1.2	0.4
Aziridine	1	30.2	30.5	30.2	0.3	0.3	0.0
Cyanogen	1	73.3	75.9	75.3	2.6	0.6	2.0
Dimethylamine	1	-4.4	-3.5	-4.5	0.9	1.0	-0.1
Ethylamine	1	-11.3	-11.6	-12.6	-0.3	0.9	-1.3
Ketene	1	-11.4	-10.7	-11.2	0.7	0.5	0.2
Oxirane	1	-12.6	-12.4	-12.7	0.2	0.4	-0.1
Acetaldehyde	1	-39.7	-39.3	-39.8	0.4	0.4	-0.1
Glyoxal	1	-50.7	-50.2	-51.0	0.5	0.8	-0.3
Ethanol	1	-56.2	-55.9	-57.0	0.3	1.1	-0.8
dimethylether	1	-44.0	-43.5	-44.5	0.5	1.1	-0.5
Thiirane	1	19.6	18.5	17.8	-1.2	0.6	-1.8
Dimethyl Sulfoxide	1	-36.2	-36.1	-36.7	0.1	0.6	-0.5

Table 2. (continued)

Ethanethiol	1		-11.1	-10.5	-11.7	0.6	1.3	-0.6
Dimethylsulfide	1		-8.9	-8.8	-9.9	0.1	1.1	-1.0
Vinyl Fluoride	1		-33.2	-33.8	-34.0	-0.6	0.2	-0.8
Ethyl Chloride	1		-26.8	-26.9	-27.2	-0.1	0.3	-0.4
Vinyl Chloride	1		8.9	5.6	5.2	-3.3	0.4	-3.7
Acrylonitrile	1		43.2	46.3	45.8	3.1	0.5	2.6
Acetone	1		-51.9	-51.3	-51.9	0.6	0.5	0.0
Acetic acid	1		-103.4	-102.7	-103.4	0.7	0.7	0.0
Acetyl Fluoride	1		-105.7	-104.6	-105.1	1.1	0.5	0.6
Acetyl Chloride	1		-58.0	-57.1	-57.7	1.0	0.7	0.3
Propyl Chloride	1		-31.5	-32.0	-32.3	-0.5	0.3	-0.8
Isopropanol	1		-65.2	-64.6	-66.1	0.6	1.5	-0.9
Methylethyl Ether	1		-51.7	-51.3	-52.7	0.4	1.4	-1.0
Trimethylamine	1		-5.7	-6.1	-6.3	-0.4	0.2	-0.6
Furan	1		-8.3	-6.3	-7.7	2.1	1.4	0.6
Thiophene	1		27.5	29.1	27.4	1.6	1.7	-0.1
Pyrrole	1		25.9	27.7	26.3	1.8	1.3	0.4
Pyridine	1		33.6	35.5	34.6	1.9	0.8	1.0
H_2	1		0.0	-0.4	-0.4	-0.4	-0.1	-0.4
HS	2		34.2	34.4	34.1	0.1	0.3	-0.1
CCH	2		135.1	136.8	137.7	1.7	-0.8	2.6
C_2H_3	2		71.6	72.6	72.1	1.0	0.5	0.5
CH_3CO	2		-2.4	-0.5	-2.2	1.9	1.7	0.2
H_2COH	2		-4.1	-3.3	-4.3	0.8	1.0	-0.2
CH_3O	2		5.0	5.1	5.0	0.1	0.0	0.0
CH_3CH_2O	2		-3.3	-0.4	-3.6	2.9	3.2	-0.3
CH_3S	2		29.8	30.0	29.1	0.2	0.9	-0.7
C_2H_5	2		28.9	29.6	29.0	0.7	0.6	0.1
$(CH_3)_2CH$	2		21.5	23.6	21.5	2.1	2.1	0.0
$(CH_3)_3C$	2		12.3	14.7	13.1	2.4	1.6	0.8
NO_2	2		7.9	11.9	9.0	4.0	2.9	1.1
			MAD G2-2			1.1	0.8	0.8
			MAD G2-2 singlets			1.1	0.7	0.8
			MAD G2 set			1.0	0.7	0.8
			MAD G2 set singlets			0.9	0.6	0.8
			MAD G2-2 non singlets			1.5	1.3	0.6
			MAX error G2 set			4.6	3.2	4.2

Table 3: ccCA-CC(2,3) Heats of Formation and MAD[a] values (kcal/mol) of the G2-1 set from experiment (expt.), GROS4-CC(2,3), and uROS4.

G2-1 Set		ccCA-CC(2,3) ΔH_f (298K), kcal/mol					
Species	Multiplicity	expt. (1)	GROS4-CC(2,3)(2)	uROS4 (3)	(2)-(1)	(2)-(3)	(3)-(1)
LiH	1	33.3	32.4	32.4	-0.9	0.0	-0.9
BeH	2	81.7	81.6	81.0	-0.1	0.5	-0.7
CH	2	142.5	142.4	142.3	-0.2	0.0	-0.2
CH_2	3	93.7	94.3	94.2	0.6	0.1	0.5
CH_2	1	102.8	102.5	102.3	-0.3	0.2	-0.5
CH_3	2	35.0	35.1	35.0	0.1	0.1	0.0
CH_4	1	-17.9	-17.4	-18.0	0.5	0.5	-0.1
NH	3	85.2	85.6	85.9	0.4	-0.2	0.7
NH_2	2	45.1	44.2	44.2	-0.9	-0.1	-0.9
NH_3	1	-11.0	-11.3	-11.6	-0.3	0.3	-0.6
OH	2	9.4	8.9	8.6	-0.5	0.2	-0.8
OH_2	1	-57.8	-58.4	-59.0	-0.6	0.6	-1.2
FH	1	-65.1	-65.7	-66.1	-0.6	0.4	-1.0
SiH_2	1	65.2	64.4	63.0	-0.8	1.4	-2.2
SiH_2	3	86.2	87.4	86.1	1.2	1.3	-0.1
SiH_3	2	47.9	48.3	46.8	0.4	1.5	-1.1
SiH_4	1	8.2	8.1	6.3	-0.1	1.8	-1.9
PH_2	2	33.1	32.0	32.4	-1.1	-0.4	-0.7
PH_3	1	1.3	1.8	1.5	0.5	0.2	0.2
SH_2	1	-4.9	-5.1	-5.9	-0.2	0.8	-1.0
ClH	1	-22.1	-22.0	-22.6	0.1	0.5	-0.5
Li_2	1	51.6	50.5	50.5	-1.1	0.0	-1.1
LiF	1	-80.1	-81.4	-82.0	-1.3	0.6	-1.9
C_2H_2	1	54.2	56.1	55.2	1.9	0.9	1.0
C_2H_4	1	12.5	13.6	12.5	1.1	1.1	0.0
C_2H_6	1	-20.1	-19.1	-20.4	1.0	1.3	-0.3

[a] The MAD values are reported between each of the three data sets as explained in the text. MAD G2-1 is the MAD for the whole G2-1 set and MAD singlet species is for all of the singlet species in the G2-1 set.

Table 3. (continued)

CN	2	104.9	106.1	108.2	1.2	-2.1	3.3
HCN	1	31.5	32.2	31.5	0.6	0.7	0.0
CO	1	-26.4	-26.0	-26.8	0.4	0.9	-0.4
HCO	2	10.0	10.8	10.0	0.8	0.7	0.0
H$_2$CO	1	-26.0	-25.4	-26.5	0.6	1.1	-0.5
H$_3$COH	1	-48.0	-47.6	-48.9	0.4	1.3	-0.9
N$_2$	1	0.0	0.1	-0.2	0.1	0.2	-0.2
H$_2$NNH$_2$	1	22.8	23.2	22.1	0.3	1.0	-0.7
NO	2	21.6	22.0	22.3	0.4	-0.4	0.7
O$_2$	3	0.0	0.6	1.1	0.6	-0.5	1.1
HOOH	1	-32.5	-31.9	-33.1	0.6	1.2	-0.6
F$_2$	1	0.0	1.4	0.9	1.4	0.5	0.9
CO$_2$	1	-94.1	-92.7	-94.0	1.4	1.3	0.1
Na$_2$	1	34.0	32.5	32.9	-1.5	-0.3	-1.1
Si$_2$	3	139.9	144.6	140.3	4.7	4.3	0.4
P$_2$	1	34.3	36.0	35.5	1.7	0.5	1.2
S$_2$	3	30.7	31.6	30.9	0.9	0.7	0.2
Cl$_2$	1	0.0	1.6	0.1	1.6	1.5	0.1
NaCl	1	-43.6	-42.5	-43.2	1.1	0.6	0.4
SiO	1	-24.6	-21.3	-23.8	3.3	2.5	0.8
SC	1	66.9	67.6	66.8	0.6	0.8	-0.1
SO	3	1.2	2.6	2.1	1.4	0.5	0.9
ClO	2	24.2	26.4	25.8	2.2	0.6	1.6
FCl	1	-13.2	-12.6	-13.6	0.6	1.0	-0.4
Si$_2$H$_6$	1	19.1	19.8	16.1	0.6	3.6	-3.0
CH$_3$Cl	1	-19.6	-18.7	-20.1	0.9	1.4	-0.5
H$_3$CSH	1	-5.5	-4.6	-6.3	0.9	1.7	-0.8
HOCl	1	-17.8	-17.6	-18.9	0.3	1.4	-1.1
SO$_2$	1	-71.0	-68.6	-70.7	2.4	2.1	0.3
			MAD G2-1		0.9	0.9	0.8
			MAD singlet species		0.9	1.0	0.8

Table 4. ccCA-CC(2,3) Heats of Formation and MAD^a values (kcal/mol) for the G2-2 set from experiment (expt.), GROS4-CC(2,3), and uROS4.

G2-2 Set		ccCA-CC(2,3) ΔH_f(298K)					
		expt.	GROS4-		(2)-	(2)-	(3)-
Species	Multiplicity	(1)	CC(2,3)(2)	uROS4(3)	(1)	(3)	(1)
BF$_3$	1	-271.4	-269.5	-269.7	1.9	0.2	1.7
BCl$_3$	1	-96.3	-92.9	-95.4	3.4	2.5	0.9
AlF$_3$	1	-289.0	-285.4	-289.9	3.6	4.6	-0.9
AlCl$_3$	1	-139.7	-134.6	-140.9	5.1	6.3	-1.2
CF$_4$	1	-223.0	-221.1	-223.1	1.9	1.9	-0.1
CCl$_4$	1	-22.9	-17.7	-22.9	5.2	5.2	0.0
COS	1	-33.1	-32.3	-34.3	0.8	2.0	-1.2
CS$_2$	1	28.0	29.7	27.0	1.7	2.7	-1.0
CF$_2$O	1	-149.1	-143.2	-144.9	5.9	1.7	4.2
SiF$_4$	1	-386.0	-381.7	-385.2	4.3	3.5	0.8
SiCl$_4$	1	-158.4	-151.2	-157.5	7.2	6.3	0.9
N$_2$O	1	19.6	20.9	19.9	1.3	1.1	0.3
ClNO	1	12.4	12.9	13.3	0.5	-0.4	0.9
NF$_3$	1	-31.6	-30.1	-31.2	1.5	1.1	0.4
PF$_3$	1	-229.1	-226.0	-227.6	3.1	1.6	1.5
O3	1	34.1	38.3	36.1	4.2	2.2	2.0
F$_2$O	1	5.9	8.0	6.9	2.1	1.0	1.0
ClF$_3$	1	-38.0	-36.4	-38.7	1.6	2.3	-0.7
C$_2$F$_4$	1	-157.4	-157.6	-160.7	-0.2	3.1	-3.3
C$_2$Cl$_4$	1	-3.0	1.2	-5.3	4.2	6.5	-2.3
CF$_3$CN	1	-118.4	-114.8	-118.1	3.6	3.4	0.3
Propyne	1	44.2	46.5	44.9	2.3	1.6	0.7
Allene	1	45.5	47.4	46.0	1.9	1.4	0.5
Cyclopropene	1	66.2	69.8	68.4	3.6	1.4	2.2

^{a} The MAD values are reported between each of the three data sets as explained in the text. The five consecutive MAD labels at the end of the table are respectively, the MAD for all species in the G2-2 set, the MAD for all singlet species in the G2-2 set, the MAD for all species in the G2/97 set, the MAD for all singlet species in the G2/97 set, and the MAD for non-singlet species in the G2-2 set. The maximum (MAX) error is given for all species in the G2/97 set.

Table 4. (continued)

Propylene	1	4.8	6.2	4.8	1.4	1.4	0.0
Cyclopropane	1	12.7	14.4	12.9	1.7	1.4	0.2
Propane	1	-25.0	-24.2	-25.5	0.8	1.2	-0.5
Butadiene	1	26.3	29.6	27.1	3.3	2.4	0.8
2-Butyne	1	34.8	38.4	36.1	3.6	2.3	1.3
Methylene Cyclopropane	1	47.9	49.1	46.8	1.2	2.3	-1.1
Bicyclobutane	1	51.9	56.4	54.1	4.5	2.3	2.2
Cyclobutene	1	37.4	41.2	39.0	3.8	2.2	1.6
Cyclobutane	1	6.8	8.6	6.5	1.8	2.1	-0.3
Isobutene	1	-4.0	-1.4	-3.9	2.6	2.5	0.1
Transbutane	1	-30.0	-28.5	-30.6	1.5	2.1	-0.6
Isobutane	1	-32.1	-30.4	-32.4	1.7	2.0	-0.3
Spiropentane	1	44.3	47.8	44.7	3.5	3.1	0.4
Benzene	1	19.7	25.2	20.8	5.5	4.4	1.1
CH_2F_2	1	-107.7	-107.0	-108.1	0.7	1.1	-0.4
CHF_3	1	-166.6	-165.1	-166.7	1.5	1.5	-0.1
CH_2Cl_2	1	-22.8	-20.6	-22.9	2.2	2.3	-0.1
$CHCl_3$	1	-24.7	-20.7	-24.4	4.0	3.7	0.3
Methylamine	1	-5.5	-5.1	-5.8	0.4	0.6	-0.3
Methyl Cyanide	1	17.7	19.7	18.1	2.0	1.6	0.4
Nitromethane	1	-17.8	-15.3	-17.8	2.5	2.5	0.0
Methyl Nitrite	1	-15.9	-13.8	-16.0	2.1	2.2	-0.1
Methyl Silane	1	-7.0	-5.6	-7.5	1.4	1.9	-0.5
Formic Acid	1	-90.5	-89.3	-91.0	1.2	1.6	-0.5
Methyl Formate	1	-85.0	-83.7	-86.1	1.3	2.4	-1.1
Acetamide	1	-57.0	-54.2	-56.6	2.8	2.3	0.4
Aziridine	1	30.2	31.6	30.2	1.4	1.4	0.0
Cyanogen	1	73.3	77.5	75.3	4.2	2.2	2.0
Dimethylamine	1	-4.4	-3.3	-4.5	1.1	1.2	-0.1
Ethylamine	1	-11.3	-11.4	-12.6	-0.1	1.2	-1.3
Ketene	1	-11.4	-9.9	-11.2	1.5	1.3	0.2
Oxirane	1	-12.6	-11.3	-12.7	1.3	1.5	-0.1
Acetaldehyde	1	-39.7	-38.3	-39.8	1.4	1.5	-0.1
Glyoxal	1	-50.7	-48.5	-51.0	2.2	2.5	-0.3
Ethanol	1	-56.2	-55.5	-57.0	0.7	1.5	-0.8
Dimethylether	1	-44.0	-43.0	-44.5	1.0	1.5	-0.5
Thiirane	1	19.6	19.9	17.8	0.3	2.1	-1.8
Dimethyl Sulfoxide	1	-36.2	-33.6	-36.7	2.6	3.0	-0.5

Table 4. (continued)

Ethanethiol	1	-11.1	-9.7	-11.7	1.5	2.1	-0.6
Dimethylsulfide	1	-9.0	-7.9	-9.9	1.1	2.0	-0.9
Vinyl Fluoride	1	-33.2	-32.7	-34.0	0.5	1.3	-0.8
Ethyl Chloride	1	-26.8	-25.4	-27.2	1.4	1.8	-0.4
Vinyl Chloride	1	8.9	7.1	5.2	-1.8	1.9	-3.7
Acrylonitrile	1	43.2	48.2	45.8	5.0	2.4	2.6
Acetone	1	-51.9	-49.6	-51.9	2.3	2.3	0.0
Acetic Acid	1	-103.4	-100.9	-103.4	2.5	2.5	0.0
Acetyl Fluoride	1	-105.7	-103.1	-105.1	2.6	1.9	0.6
Acetyl Chloride	1	-58.0	-55.0	-57.7	3.0	2.8	0.3
Propyl Chloride	1	-31.5	-29.8	-32.3	1.7	2.5	-0.8
Isopropanol	1	-65.2	-63.8	-66.1	1.4	2.3	-0.9
Methylethyl Ether	1	-51.7	-50.5	-52.7	1.2	2.2	-1.0
Trimethylamine	1	-5.7	-4.4	-6.3	1.3	1.9	-0.6
Furan	1	-8.3	-4.3	-7.7	4.0	3.4	0.6
Thiophene	1	27.5	31.6	27.4	4.1	4.2	-0.1
Pyrrole	1	25.9	29.5	26.3	3.6	3.2	0.4
Pyridine	1	33.6	39.0	34.6	5.4	4.4	1.0
H_2	1	0.0	-0.4	-0.4	-0.4	-0.1	-0.4
HS	2	34.2	34.3	34.1	0.1	0.2	-0.1
CCH	2	135.1	135.9	137.7	0.8	-1.8	2.6
C_2H_3	2	71.6	72.3	72.1	0.7	0.2	0.5
CH_3CO	2	-2.4	-0.8	-2.2	1.6	1.4	0.2
H_2COH	2	-4.1	-3.3	-4.3	0.8	1.0	-0.2
CH_3O	2	5.0	4.8	5.0	-0.2	-0.3	0.0
CH_3CH_2O	2	-3.3	-0.7	-3.6	2.6	2.9	-0.3
CH_3S	2	29.8	29.8	29.1	0.0	0.7	-0.7
C_2H_5	2	28.9	29.4	29.0	0.5	0.4	0.1
$(CH_3)_2CH$	2	21.5	22.6	21.5	1.1	1.1	0.0
$(CH_3)_3C$	2	12.3	12.8	13.1	0.5	-0.2	0.8
NO_2	2	7.9	9.6	9.0	1.7	0.6	1.1

MAD G2-2	2.2	2.1	0.8
MAD G2-2 singlets	2.4	2.3	0.8
MAD G2 set	1.7	1.7	0.8
MAD G2 set singlets	1.9	1.9	0.8
MAD G2-2 non-singlets	0.9	0.9	0.6
MAX error G2 set	7.2	6.5	4.2

Table 5. Electronic energies corrected for zero-point vibrational energies, i.e., enthalpies at 0K[a] (kcal/mol), characterizing the con_TS and dis_TS transition states, g-but intermediate, gt_TS transition state, and t-but final product at several levels of theory. All energies are relative to bicbut reactant.

Method	con_TS	dis_TS	g-but	gt_TS	t-but
CASSCF(10,10)/cc-pVDZ	38.5	52.0	-39.7	-37.2	-42.6
CCSD(T)/ccpVTZ// CASSCF(10,10)/cc-pVDZ	40.4	21.8	-25.1	-22.3	-28.0
CR-CC(2,3)/ccpVTZ// CASSCF(10,10)/cc-pVDZ	41.1	66.1	-24.9	-22.1	-27.9
ccCA-S4// CASSCF(10,10)/cc-pVDZ	40.4	21.5	-24.8	-21.9	-27.8
ccCA-CC(2,3)// CASSCF(10,10)/cc-pVDZ	41.1	65.7	-24.6	-21.8	-27.7
% diradical character[b]	24	90	9	8	9
Experiment	40.6±2.5[c]				-25.9±0.4[d]

[a] The ccCA values are based on heats formation at 0K. The CASSCF and CR-CC(2,3) values are taken from Ref. 50. The A//B notation indicates an energy calculated at level A using the geometry determined at level B.

[b] Obtained using CASSCF(10,10)/cc-pVDZ; bicbut has a diradical character of 4%.

[c] Ref. 73.

[d] Reaction enthalpy at 298K based on the enthalpies of formation of bicyclo[1.1.0]butane and buta-1,3-diene in Ref. 74.

Chapter 5

General Conclusions

In Chapter 2, a reactive force field was used to obtain the lowest energy structures of catalytic mesoporous silica nanoparticle (MSN) systems, which have been characterized using mobility data from solid-state ^{13}C NMR. The reactive force field used is ReaxFF, which is a bond order dependent force field that can treat silica systems, such as those found in MSN systems. ReaxFF was implemented into and used via the GAMESS computational package. The model used to represent the MSN systems contain a four-layer silica (4L-MSN) slab with catalytic substituents attached to the surface of the slab. These catalytic substituents were 3-cyanopropyl- (CP), N-(2-aminoethyl)-3-aminopropyl- (AAP), and N-[N-(2-aminoethyl)-2-aminoethyl]-3-aminopropyl- (AEP). From ^{29}Si DPMAS (direct polarization magic angle spinning) solid-state NMR experiment, it was determined that a single catalytic substituent exists on a 1nm^2 surface area of the 4L-MSN and the ReaxFF MSN structures were modeled as such. Also, using ^{13}C{^{1}H} CPMAS (cross polarization magic angle spinning) solid-state NMR indicated that AAP and AEP show very low mobility on the surface of the 4L-MSN, while CP shows two orders of magnitude higher mobility of the surface of the 4L-MSN. It was theorized that low mobilities for AAP and AEP reflect that these catalytic substituents exist primarily in the prone (or flat) position across the surface of the 4L-MSN due to the amine groups on AAP and AEP interacting with the many hydroxyl groups on the 4L-MSN surface. However, CP has little interaction with the silica surface and is reflected in a higher mobility. In order to test this theory, all catalytic substituents were placed on the 4L-MSN surfaces in two configurations: upright and prone. ReaxFF was used to obtain the relative energies of these configurations for each of the catalytic substituents,

since this force field can treat silica systems well and has low computational cost. It was shown that all three MSN systems favored the prone position, however, the AAP- and AEP-MSN systems had the highest relative energies between their respective prone and upright conformers. The CP-MSN system is predicted to have the lowest relative energy between the prone and upright conformers. These ReaxFF based results indicate that AAP, and AEP are more strongly bound to the MSN surface, while CP is weakly attached to the MSN surface. Thus, the ReaxFF results correlate with the mobility data from solid state NMR and provide some understanding of the behavior and structure of catalysts within MSNs.

In Chapter 3, a new Quantum Mechanical/Molecular Mechanical (QM/MM) method was introduced and used to reproduce structural data on silica clusters (Si_9H_{12} and $Si_{38}H_{36}$), as well as reproduce the similar trends in the MSN systems used in Chapter 2. The new QM/MM method is a Surface Integrated Molecular Orbital/Molecular Mechanics method called SIMOMM-Rx. In SIMOMM-Rx, the MM method used is ReaxFF, which replaces the Tinker MM method used in the original SIMOMM. The SIMOMM-Rx method was tested on the Si_9H_{12} and $Si_{38}H_{36}$ clusters against other QM/MM techniques (Weiner and SIMOMM-Tinker), as well as pure quantum mechanical (GVB-PP(1)) and molecular mechanics methods (Tinker). The GVB-PP(1) method was used as the QM method in SIMOMM-Rx. The dimer bond lengths in these silicon clusters were found to only differ by ±0.03 Å between SIMOMM-Rx and the other QM/MM methods. The SIMOMM-Rx method was validated using the calibration with the silicon clusters, which are harder to characterize than silica systems since they exhibit multi-reference behavior. The SIMOMM-Rx method was then used to predict the ReaxFF trends for the prone vs. up relative energies for AAP, AEP, and CP on the 4L-MSN surface. However, these relative energies as a function of substituent

(AAP vs. AEP vs. CP) are considerably attenuated in comparison with the previously reported ReaxFF calculations. The calculated SIMOMM-Rx relative energies are nearly independent of the QM method used. The QM methods used were DFT/B3LYP, DFT/PBE0, and MP2, all with the 6-311+G(d) basis set. These QM methods used within SIMOMM-Rx predict very similar prone vs. up relative energies. Of the three substituents, CP results in the smallest prone vs up relative energy, in agreement with ReaxFF, but the variation with substituents is small. Thus, SIMOMM-Rx has been shown to provide a useful tool for the study of systems like silica nanoporous materials, for which fully QM calculations are computationally challenging. Further extensions of SIMOMM-Rx will include adding a method that will enable the study of solvent effects on surface science.

In Chapter 4, a composite method was used to produce the correct reaction pathway of the pericyclic rearrangement of bicyclo[1.1.0]butane. A new composite method was developed and implemented into GAMESS in order to properly account for the diradical species found within this pericyclic rearrangement and to obtain the correct reaction pathway. This new method is called ccCA-CC(2,3) and it incorporates the correlation consistent Composite Approach (ccCA) method and the completely renormalized coupled cluster theory with singles, doubles, and non-iterative triples excitations (CR-CC(2,3). CR-CC(2,3) has the advantage that it can treat diradicals and this advantage was translated into ccCA-S4. The ccCA-CC(2,3) method was first calibrated against ccCA-S4 heats of formation values using the G2/97 set of 148 species.

It was demonstrated that the CR-CC(2,3) method is an accurate and computationally cost effective option for the calculation of heats of formation in the G2/97 set, both within the ccCA-CC(2,3) approach, in which the CR-CC(2,3) method is used for closed- as well

open-shell species, and within the modified ccCA-S4 protocol, in which CCSD(T) is used for closed-shell systems and CR-CC(2,3) is a substitute for CCSD(T) in calculations involving open-shell species. For open shell species and diradicals, which can frequently be problematic for composite methods, the ccCA-CC(2,3) mean absolute deviation (MAD) relative to experiment, is smaller than the MAD for ccCA-S4. These results make the new method competitive with the successful and well-established ccCA-S4 approach, particularly for chemical systems that contain open shells species and diradicals. The ccCA-CC(2,3) method was then used to successfully obtain the correct reaction pathway involved in the pericyclic rearrangement of bicyclo[1.1.0]butane to trans-1,3-butadiene. It is notable that the successful use of the CR-CC(2,3) method within ccCA-CC(2,3) suggests that many composite methods that use CCSD(T) may be improved in their treatment of open shell species and diradicals by using CR-CC(2,3) instead of CCSD(T).